Maritime Informatics & Robotics – Maritime 2023
3 Jul 2023 to 12 Jul 2023, Syros

Maritime Analytics & Forecasting

Eva Chondrodima, Yannis Theodoridis
Data Science Lab., University of Piraeus

Jul 10, 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

DATA STORIES
Data Science Lab. @ University of Piraeus

UNIVERSITY OF PIRAEUS
RESEARCH CENTER

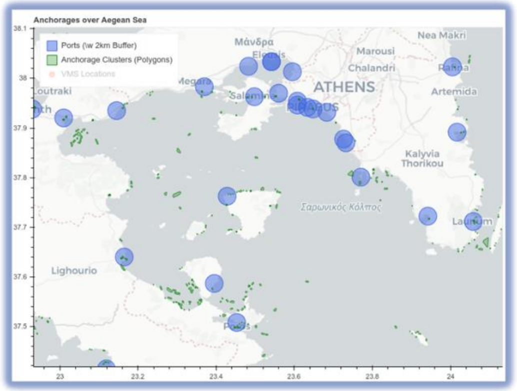
The **University of Piraeus Research Center (UPRC)** facilitates the research activities of university members in different programmes and initiatives. In this context, the Department of Informatics (through UPRC) has been actively involved in a significant number of (i) EU funded R&D projects, (ii) National projects funded by the Greek Ministry of Development and the General Secretariat of Research and Technology, and (iii) Projects developed in collaboration with enterprises (both international and national).

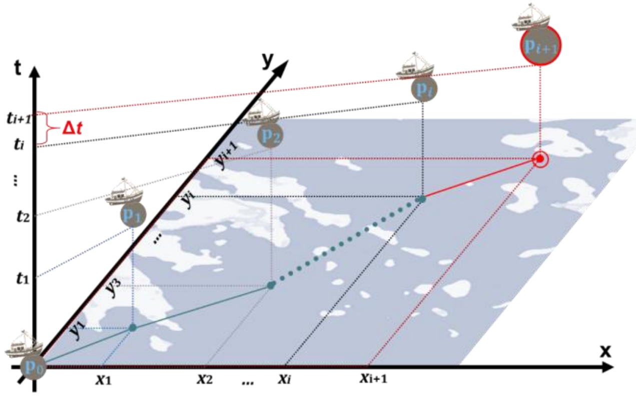
Research statement

The Data Science Lab @ Univ. Piraeus (est. 2015), aims to advance research on a wide range of Data Science topics, including **big data management**, **statistics and data analytics**, **machine learning**, **semantic integration**, with particular interest in **mobility data**.

Analytics & Forecasting Methods in the Martitime Domain

- Introduction – Getting to know maritime data
- Pre-processing methods for maritime data
- Artificial Intelligence
- Real World Problems – Applications
  - Vessel Location/Route Forecasting
    - Fishing Vessels Activity Prediction
  - Vessel Traffic Flow Forecasting
  - Vessel Collision Risk Assessment





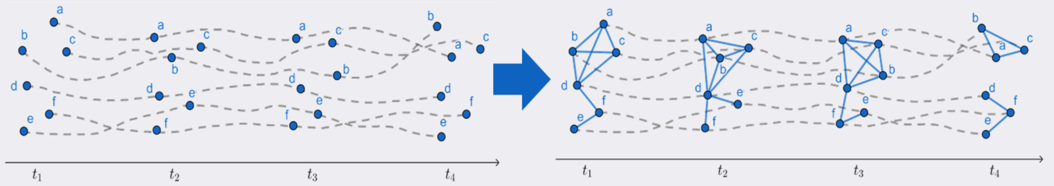


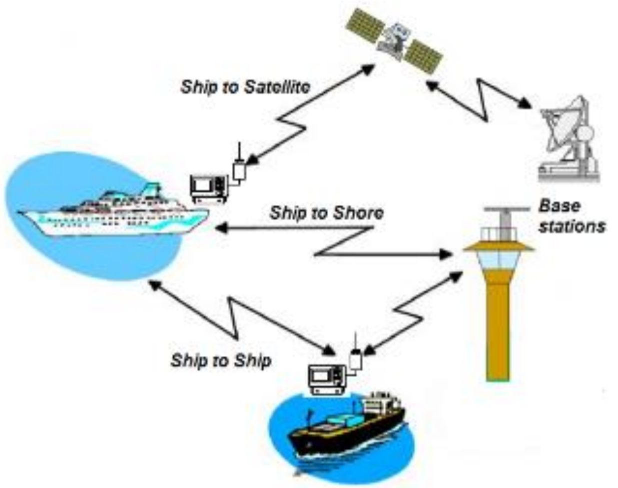
Fig.1: Example of six objects moving at four consecutive time points and the respective connectivity graphs.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

DATA STORIES
Data Science Lab. @ University of Piraeus

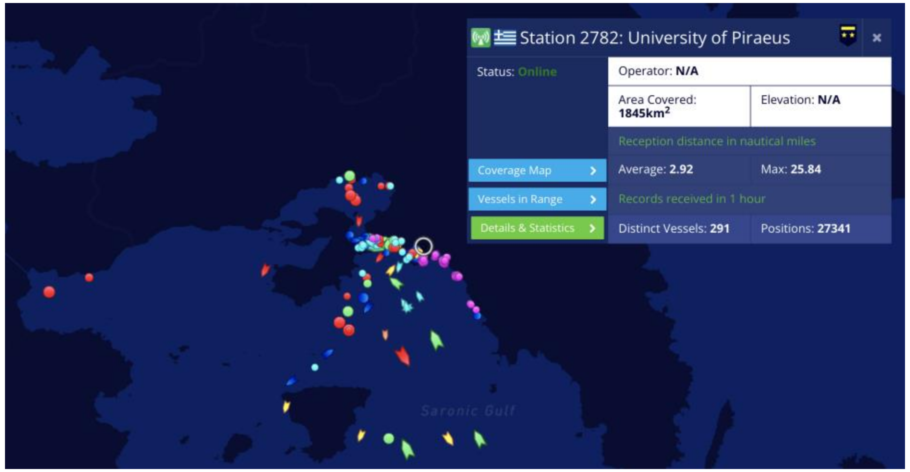
Introduction –
Getting to know maritime data

Examples of maritime datasets



AIS (Automatic Identification System): is a collaborative, self-reporting, short-range, coastal **tracking system** that **allows vessels to broadcast their identification information**, characteristics and destination, along with other information originating from on-board devices and sensors, such as location, speed and heading.

- >250,000 vessels tracked daily (source: marinetraffic.com)
- AIS signal transmitted: every 2 to 10 sec depending on speed while underway; every 3 min while at anchor



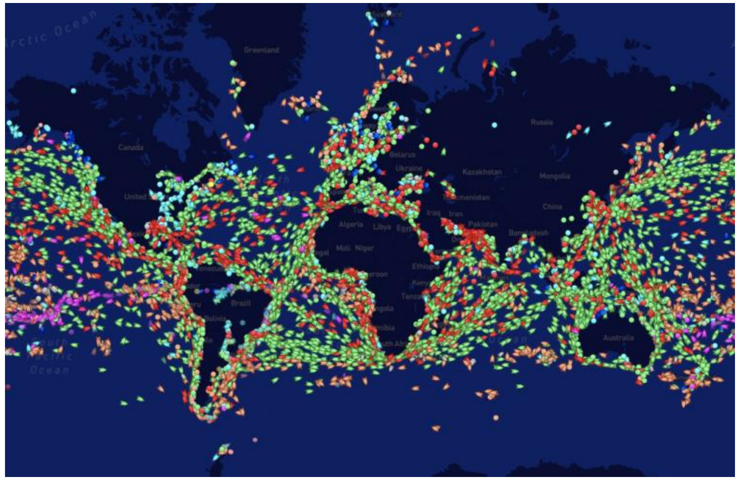


image source: marinetraffic.com

- top: global snapshot on May 26th, 2022; vessel colors correspond to different vessel types (e.g., cargo is green, tanker is red)
- left: vessels tracked by the Univ. Piraeus' AIS station

AIS signal example

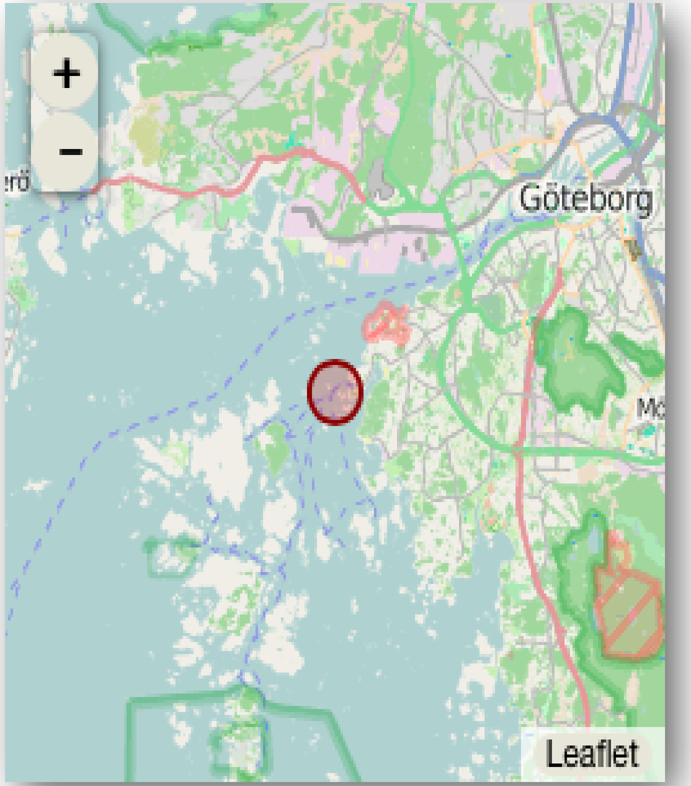
```
!AIVDM,1,1,,A,13u?etPv2;0n:dDPwUM1U1Cb069D,0*24

Message sent (UTC) : 17:21:53
MMSI               : 265547250
Latitude           : 57.660353°
Longitude          : 11.832977°
Speed              : 13.9 knots
Heading            : 41°
Course over ground : 40°
Rate of turn       : -2°/min
Navigational status: 0
Nearest place      : Styrsö, Sweden

RL.SE AIVDM/AIVDO Decoder, build #213
```

MMSI (Maritime Mobile Service Identity) is a unique 9 digit number for identifying a ship.

source: http://rl.se/aivdm



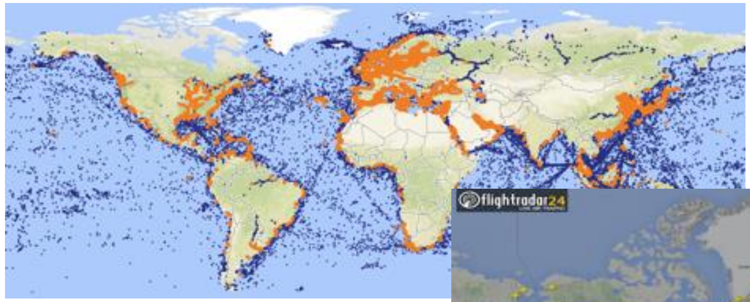
Use AIS data to improve maritime transport systems

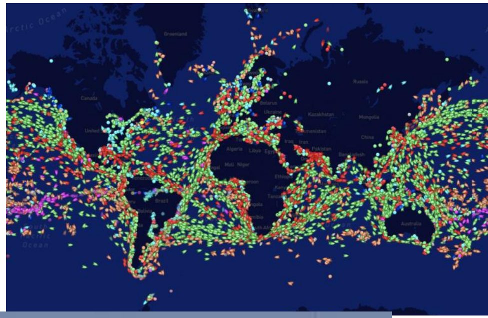
- Nowadays, vessel's movement information has become increasingly available due to the vast spread of AIS data. In order to use AIS data to improve maritime transport systems, we need to:
  - **Extract knowledge** from AIS data
  - **Learn** from AIS data
  - **Model vessel movement behaviour** based on AIS data

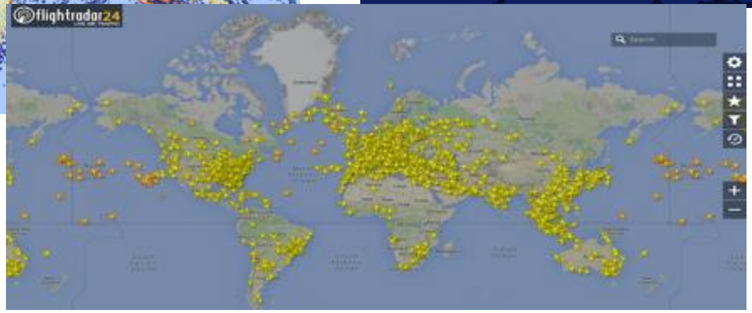
Non-trivial tasks

- Examples:
  - Find vessels that move together (for long time)
  - Find the most typical among vessels' routes as well as the outliers
  - Find the most crowded areas or routes
  - Forecast the anticipated route of a vessel or traffic in an area, etc.

Big Data problem!



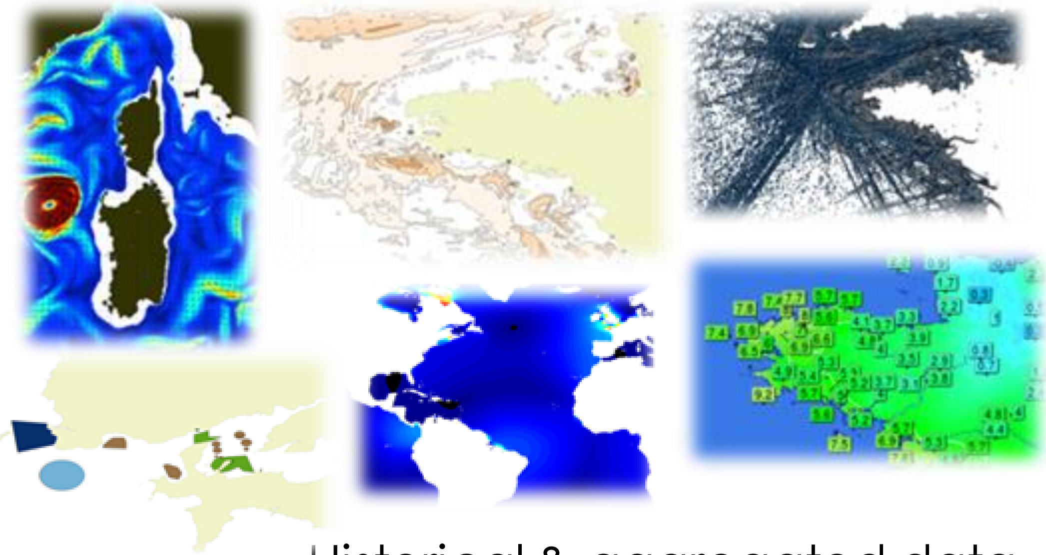




Big Data challenges

Volume
Velocity

12K distinct ships/day, 200M AIS signals/month in EU waters



Variety

Historical & aggregated data, geographical & environmental data, contextual data, etc.

Veracity

Noisy and error-prone data due to receivers limited coverage, positioning devices switch-off

Image source: (Claramunt et al. 2017)

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

DATA STORIES
Data Science Lab. @ University of Piraeus

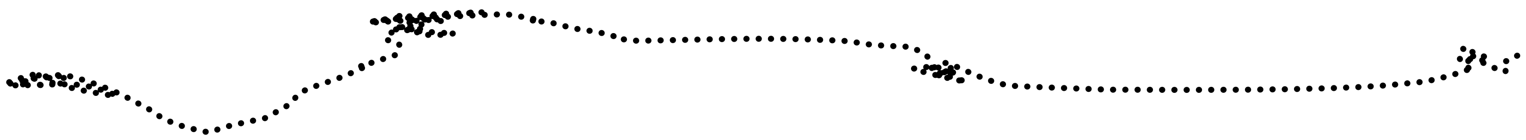
Pre-processing methods for maritime data

From GPS/AIS data to trajectories

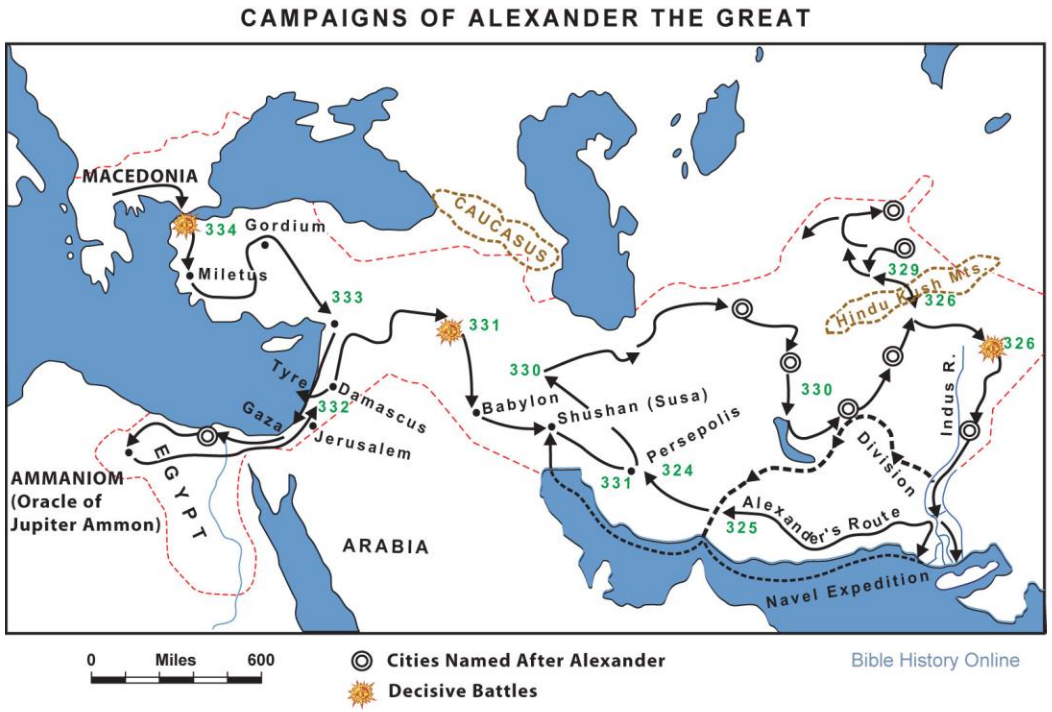
A **trajectory** is a model for a motion path of a moving object (vessel, human, animal, robot, …)

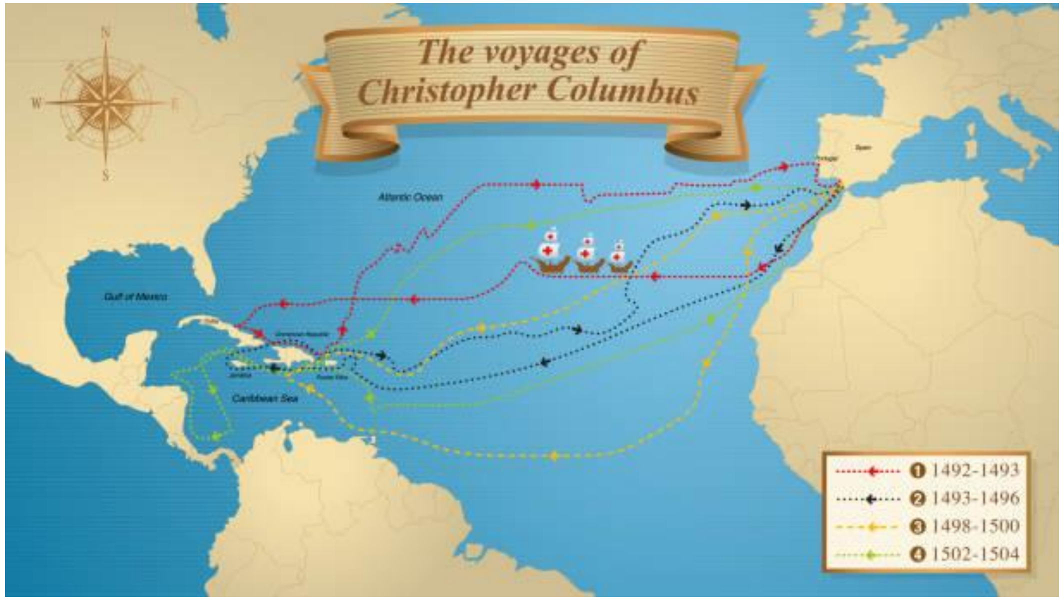
- (due to discretization) a sequence of **sampled time-stamped locations** (p_i, t_i) where:
  - p_i is a 2D or 3D point, (x_i, y_i) or (x_i, y_i, z_i) resp., and
  - t_i is the recording timestamp of p_i

$$T = \{ \langle p_1, t_1 \rangle, \langle p_2, t_2 \rangle, \ldots, \langle p_n, t_n \rangle \}$$



Popular trajectory examples

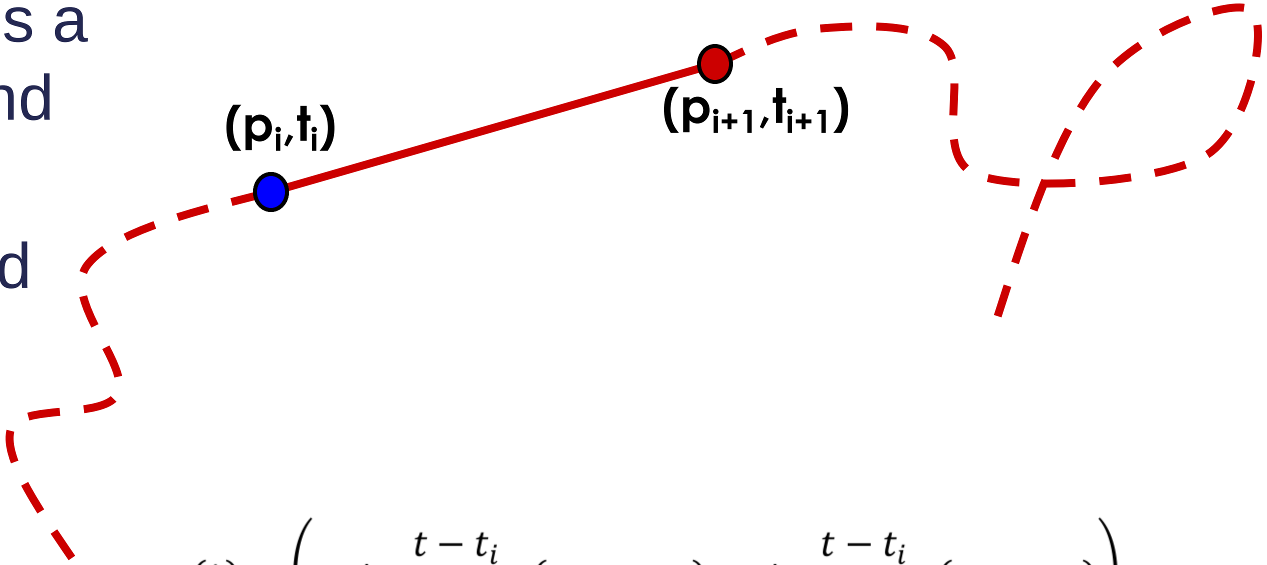




Trajectory representation

A common representation of a trajectory is a **3D/4D polyline** whose vertices correspond to time-stamped locations (p_i, t_i)

- Usually, **linear interpolation** is assumed between (p_i, t_i) and (p_{i+1}, t_{i+1})



$$p(t) = \left(x_i + \frac{t - t_i}{t_{i+1} - t_i}(x_{i+1} - x_i), y_i + \frac{t - t_i}{t_{i+1} - t_i}(y_{i+1} - y_i)\right)$$

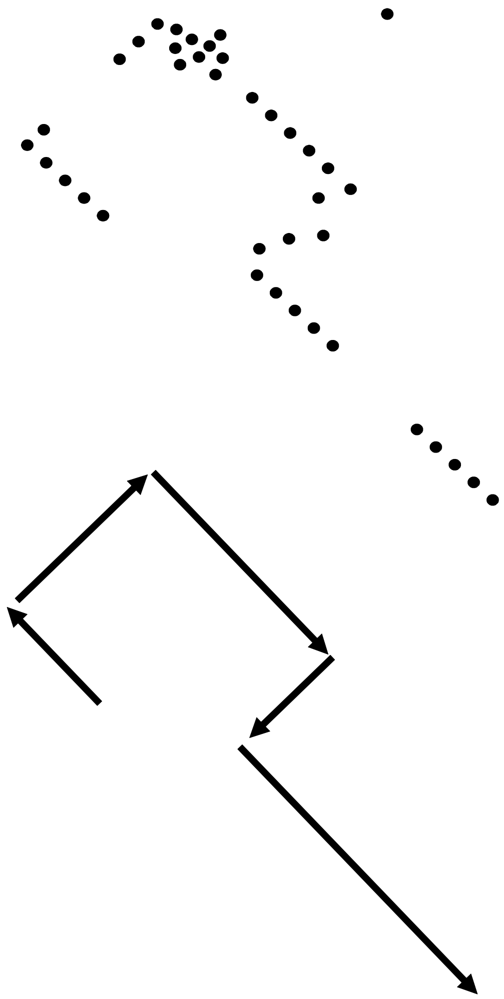
Notes:

1. Reasonable assumption only when sampling is dense
2. Does not obey the physical rules (why?)
   … but don't care (why?)

Acquiring Trajectories from Raw Data

The problem:

From raw data, i.e., successive time-stamped locations …

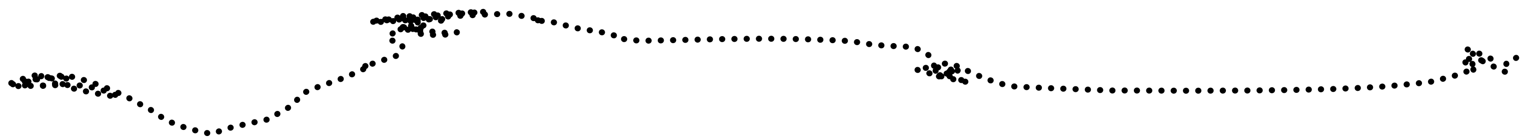


… **to meaningful trajectories**, i.e., continuous development of movement

Data pre-processing

Definition: **preparing data for analytics purposes**

$$T = \{ <p_1, t_1>, <p_2, t_2>, \ldots, <p_n, t_n> \}$$



Data pre-processing includes:

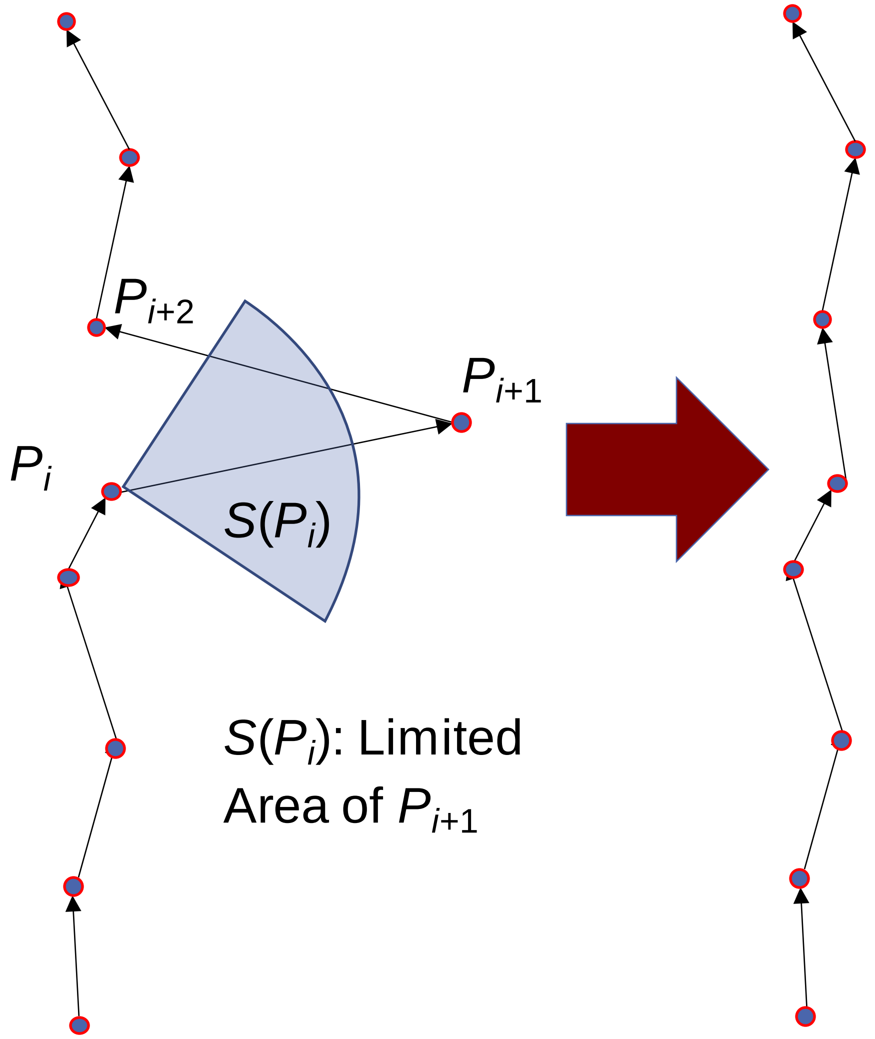
- **Cleansing**
  - noise removal
  - smoothing, etc.
- **Transformation**
  - trajectory segmentation
  - trajectory simplification, etc.
- **Enrichment**
  - semantic annotation
  - data fusion, etc.
- **etc.**

AIS Data Cleansing: Erroneous recordings - noise

Noise corresponds to values that are 'impossible' to appear

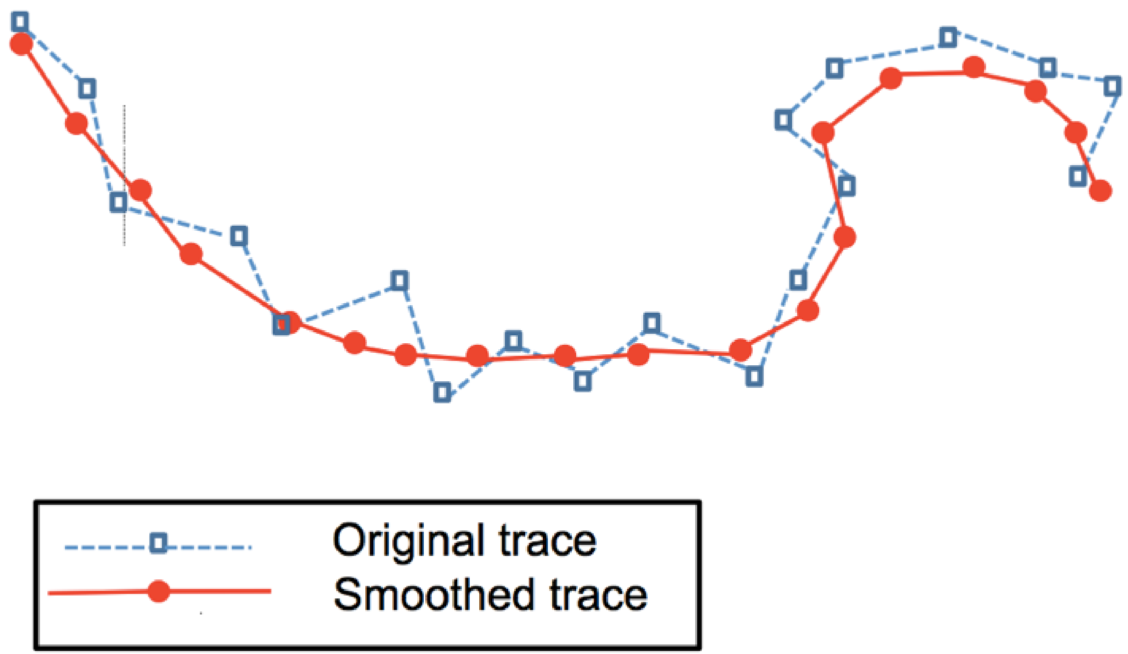
Can be detected and removed using appropriate filters

- e.g., maximum speed
  - The AIS specification for SOG (Speed over ground) shows that 102.3 knots is reported when the vessel speed is unavailable



AIS Data Cleansing: Erroneous recordings - random errors

- **Random errors** correspond to ‘possible’ values that appear to be small deviations from actual ones
- Can be smoothed using a plethora of statistical methods
  - e.g., least squares spline approximation (de Boor, 1978)

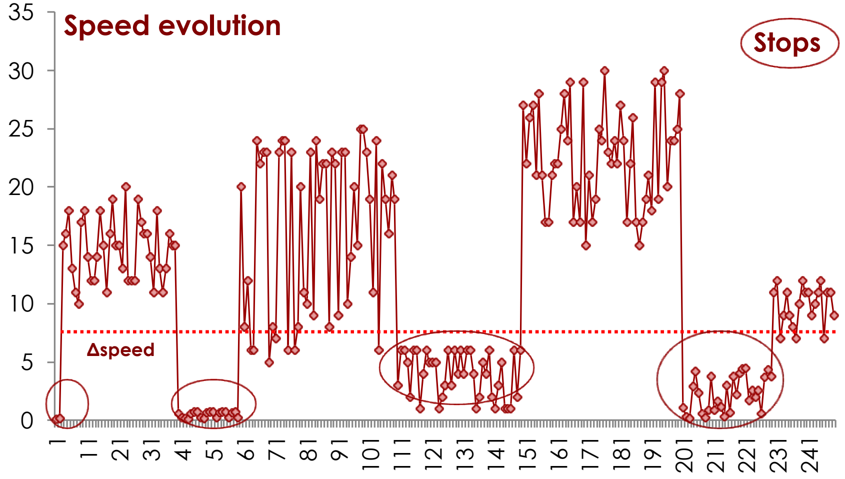


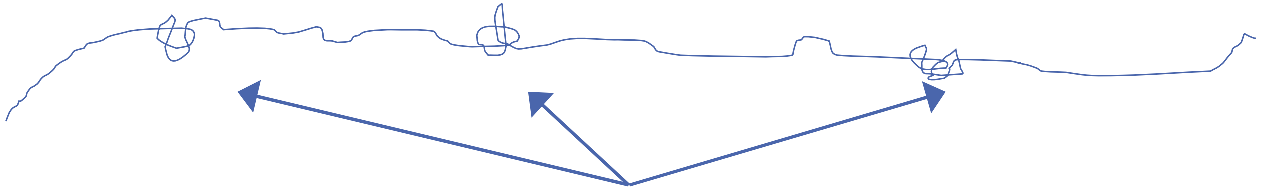
AIS Data Transformation: Trajectory segmentation

Goal: **Segment sequences of points**
in homogeneous sub-sequences

Various approaches - Segmentation via:

- raw (spatial / temporal) gap
- stop discovery
- prior knowledge (e.g. arrival at ports)
- etc.





stops

AIS Data Transformation: Trajectory simplification

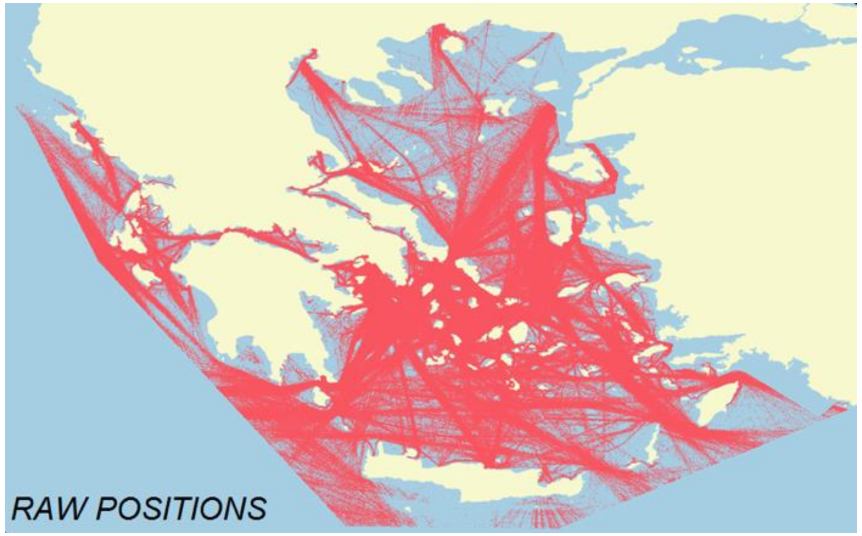
The need for simplification: efficiency in storage, processing time, etc.

- Actually, simplification **is a form of data compression**

Goal: maintain the original ‘signature’ as much as possible by keeping a set of **critical points** only

Approaches

- Offline, i.e., multi-pass, vs.
- Online, i.e., 1-pass



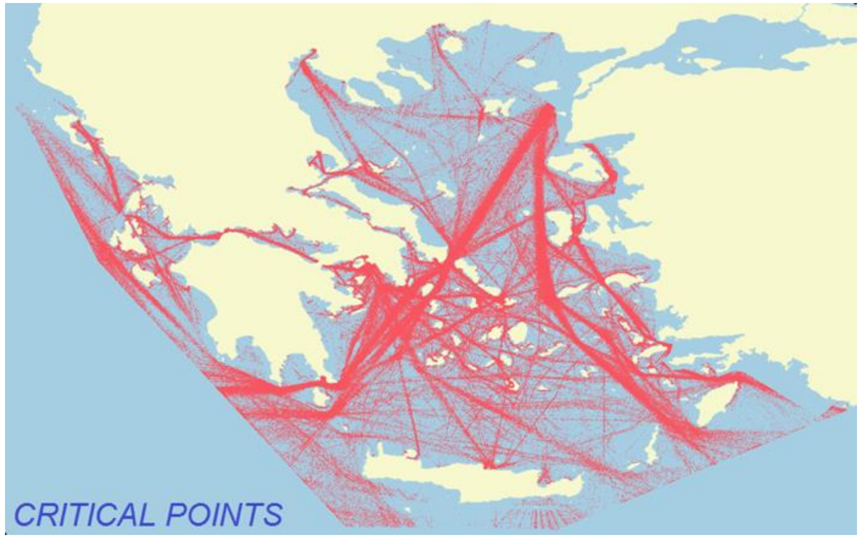


image source: aminess.eu

AIS Data Transformation: Trajectory simplification - Offline

Offline approaches:

- top-down vs. bottom-up vs. sliding window vs. opening window

e.g., **Synchronous Euclidean Distance – SED** (Meratnia & de By, 2004)

- Adapts the popular **Douglas & Peucker polyline simplification** (1973) to the mobility domain

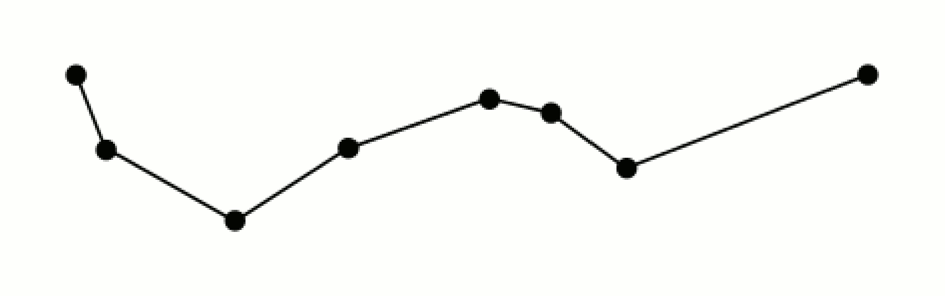
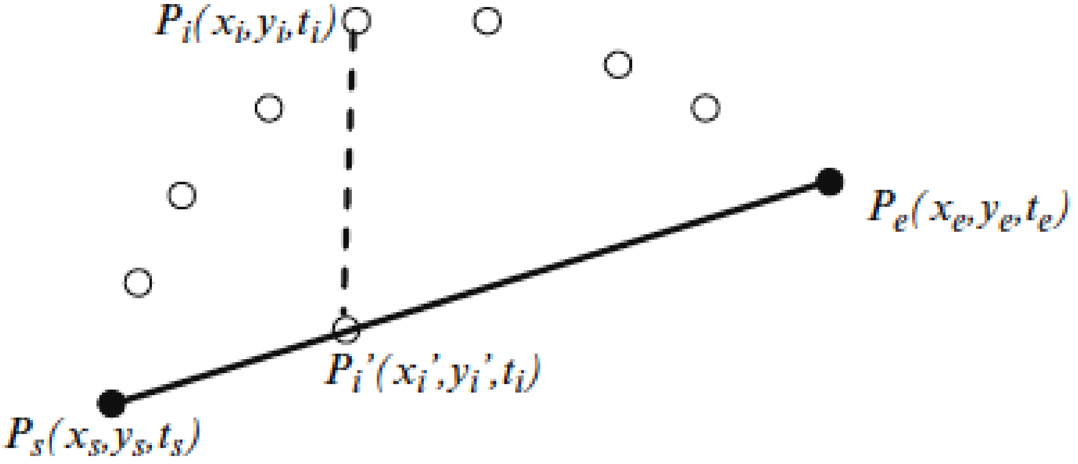


image source: https://commons.wikimedia.org/wiki/File:Douglas-Peucker_animated.gif



AIS Data Transformation: Trajectory simplification - Online

Online approaches, e.g., **Trajectory Synopses** (Patroumpas et al. 2015; 2017)

Maintains a **velocity vector** per moving object in order to detect **instantaneous events**

- stop; change in velocity vector; etc.

Tradeoff: degree of compression vs. quality of approximation

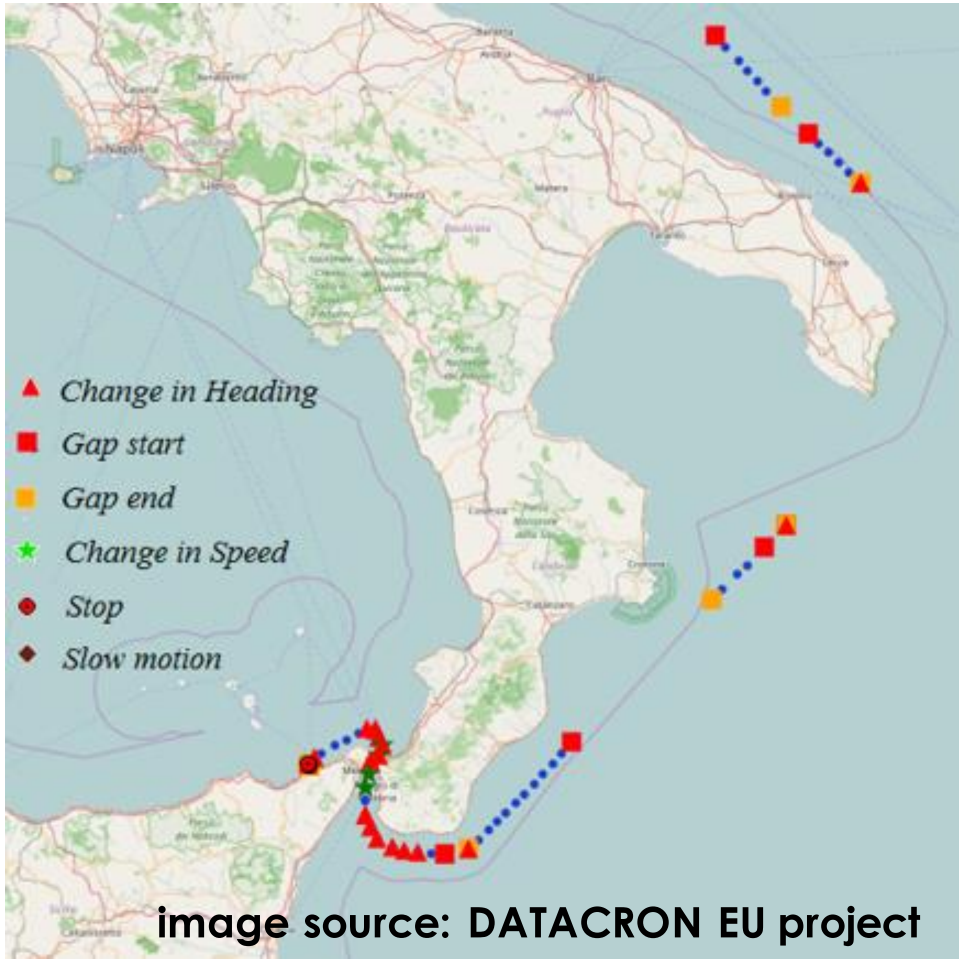


image source: DATACRON EU project

Open source: https://github.com/DataStories-UniPi/Trajectory-Synopses-Generator

Patroumpas K., et al. (2020) **Trajectory Detection and Summarization over Surveillance Data Streams**. Big Data Analytics for Time-Critical Mobility Forecasting

Artificial Intelligence (AI)

Artificial Intelligence (AI)

- The term "artificial intelligence" was first coined by **John McCarthy** (1956)
- Artificial Intelligence (AI) is the part of **computer science** concerned with designing **intelligent computer systems**, that is, systems that exhibit characteristics we associate with **intelligence in human behavior** – understanding language, learning, reasoning, solving problems, and so on.” - (Barr & Feigenbaum, 1981)
- **Intelligent behavior** involves **perception**, **reasoning**, **learning**, **communicating** and **action** in complex environments (Nilsson 1998)
- **Alan Turing** proposed in 1950 the **Turing test**, to determine whether or not a computer demonstrates intelligent behaviour.

Computational Intelligence

Nowadays, the most common way to approach AI is through the use of the so-called **Computational Intelligence (CI)** methods.

Computational Intelligence (CI) …

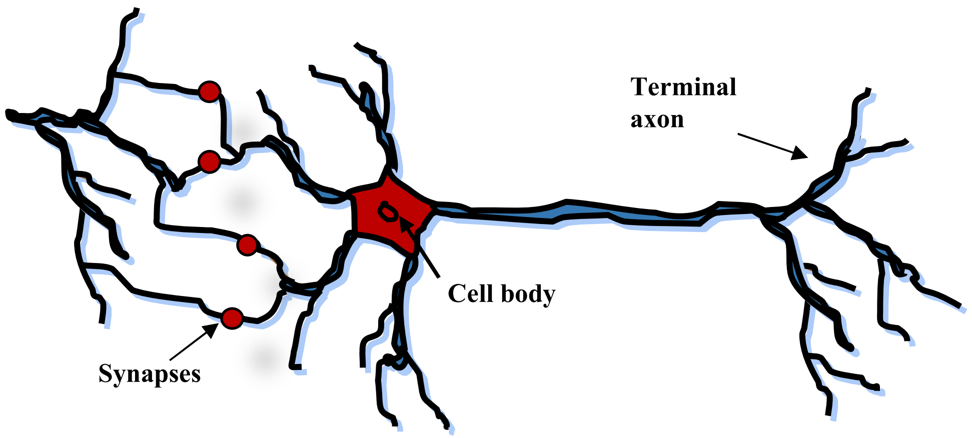
- … concept was first used in 1990 by the IEEE Neural Networks Council
- … is based on **soft computing** methods: work by aggregating data to partial truths (*much closer to the way the human brain works*)
- … is (according to **Bezdek**, 1994) a subset of AI.
- … is (according to **IEEE CIS**) the theory, design, application and development of biologically and linguistically motivated computational paradigms. Traditionally the three main pillars of CI have been **Neural Networks**, **Fuzzy Systems** and **Evolutionary Computation**.
- … is considered to encompasses **Machine Learning (ML)**, which is a subset of AI that focuses on the development of algorithms and statistical models that enable computers to learn from data, rather than relying on explicit instructions.

IEEE Computational Intelligence Society

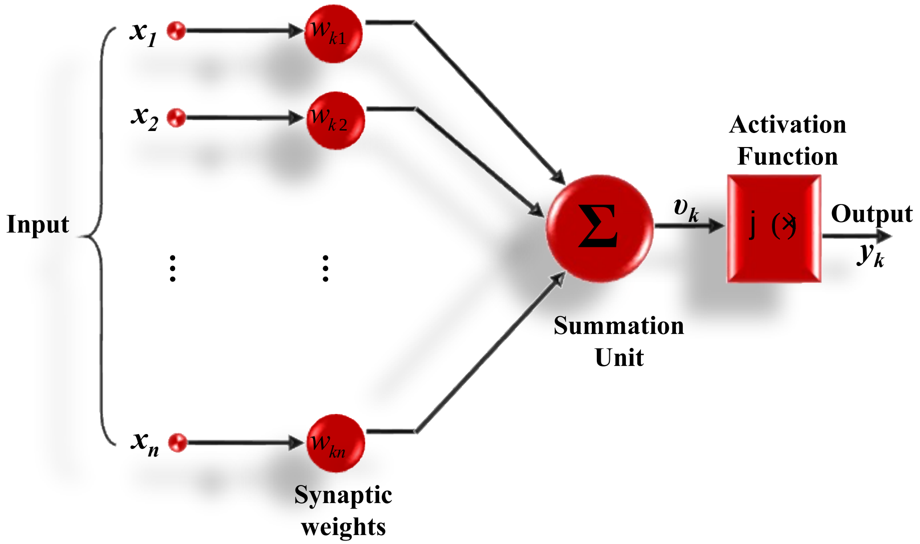
Neural Networks

- Neural networks (NNs) are a set of powerful mathematical tools that simulate the way that the human brain deals with information and the procedure of learning.
- NNs have the ability to identify and learn highly complex and nonlinear relationships from input-output data only, without the use of first principle equations describing the system.

BIOLOGICAL NEURON



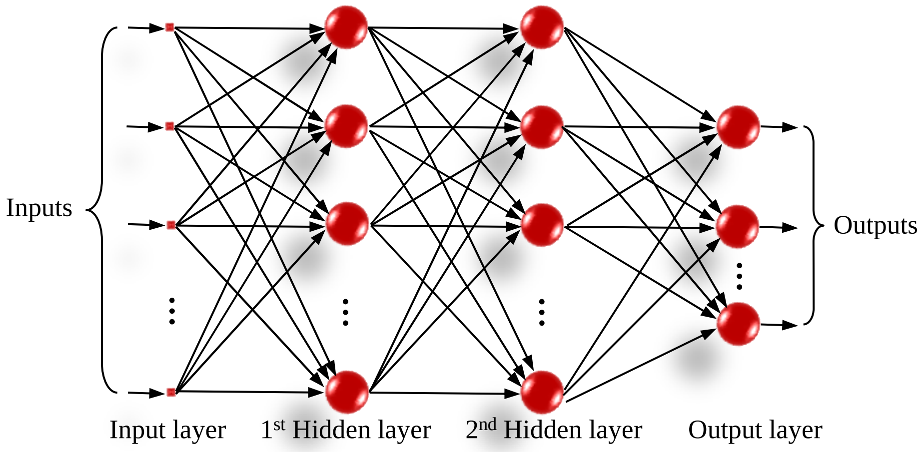
ARTIFICIAL NEURON



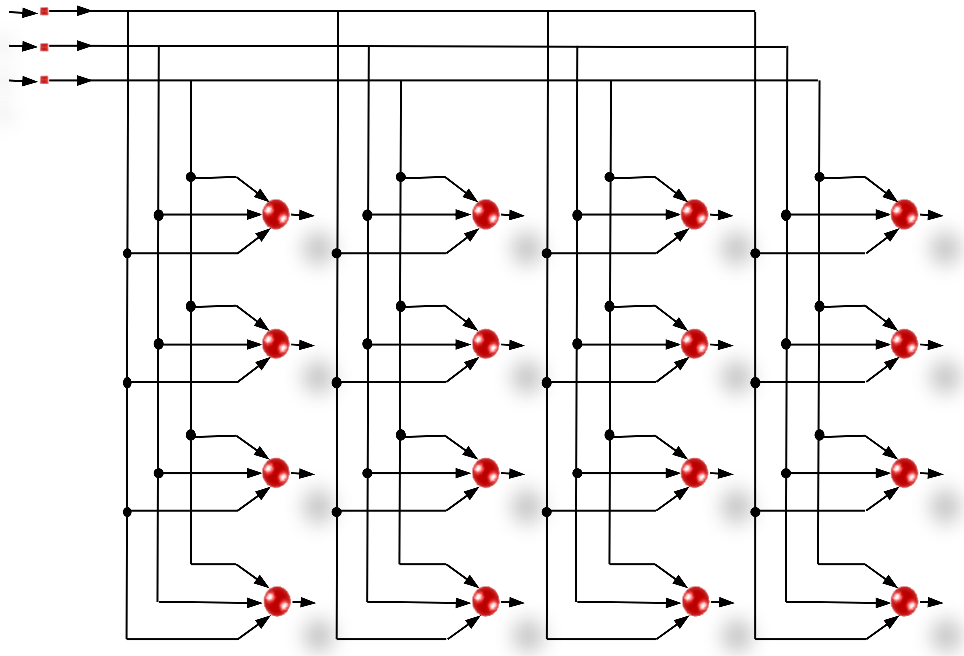
Neural Networks Architectures

- NN architecture is based on the structure and function of the biological neural network.
- Similar to neurons in the brain, NN also consists of neurons which are arranged in various layers.

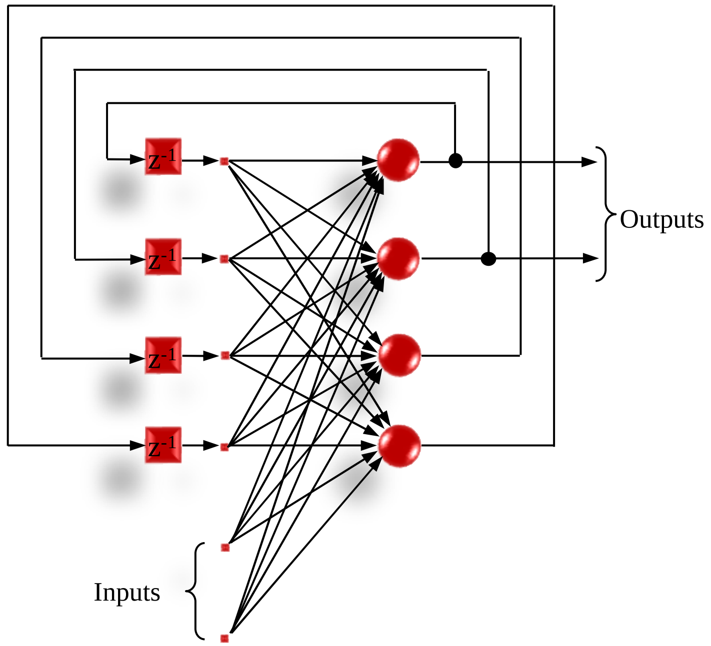
Multi-Layer Perceptrons (Feedforward networks)



Kohonen Networks (Self-Organizing Maps)



Recurrent Neural Networks

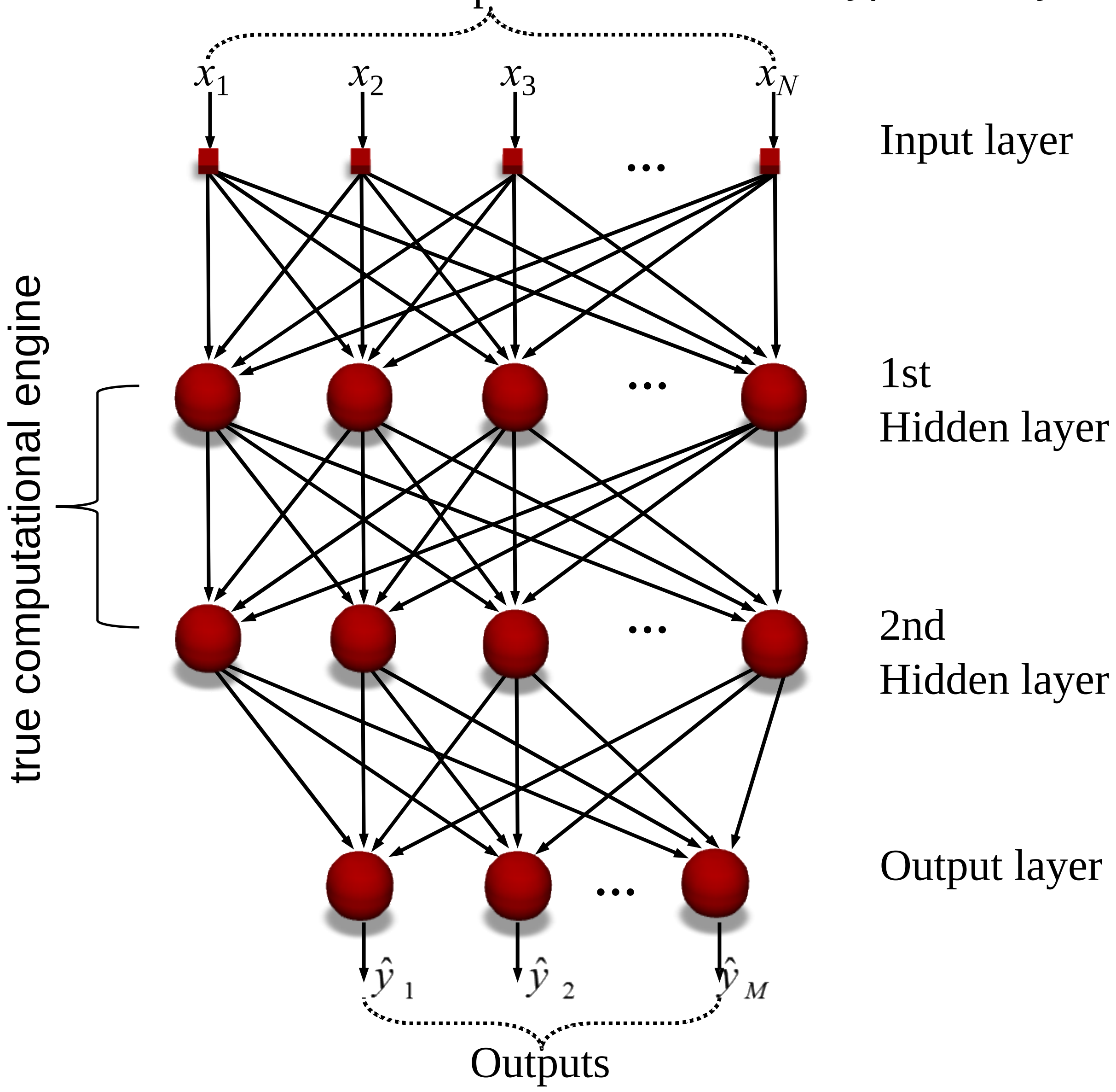


Multi-Layer Perceptrons

Multilayer Perceptron falls under the category of **feedforward algorithms**, because inputs are combined with the initial weights in a weighted sum and subjected to the activation function, just like in the Perceptron.

The data flows in the forward direction from input to output layer. Each layer is feeding the next one with the result of their computation, their internal representation of the data. This goes all the way through the hidden layers to the output layer.

3 types of layers:



Neural Networks Training

- The **goal** of a neural network is to learn **how to map input** examples **to output** examples.

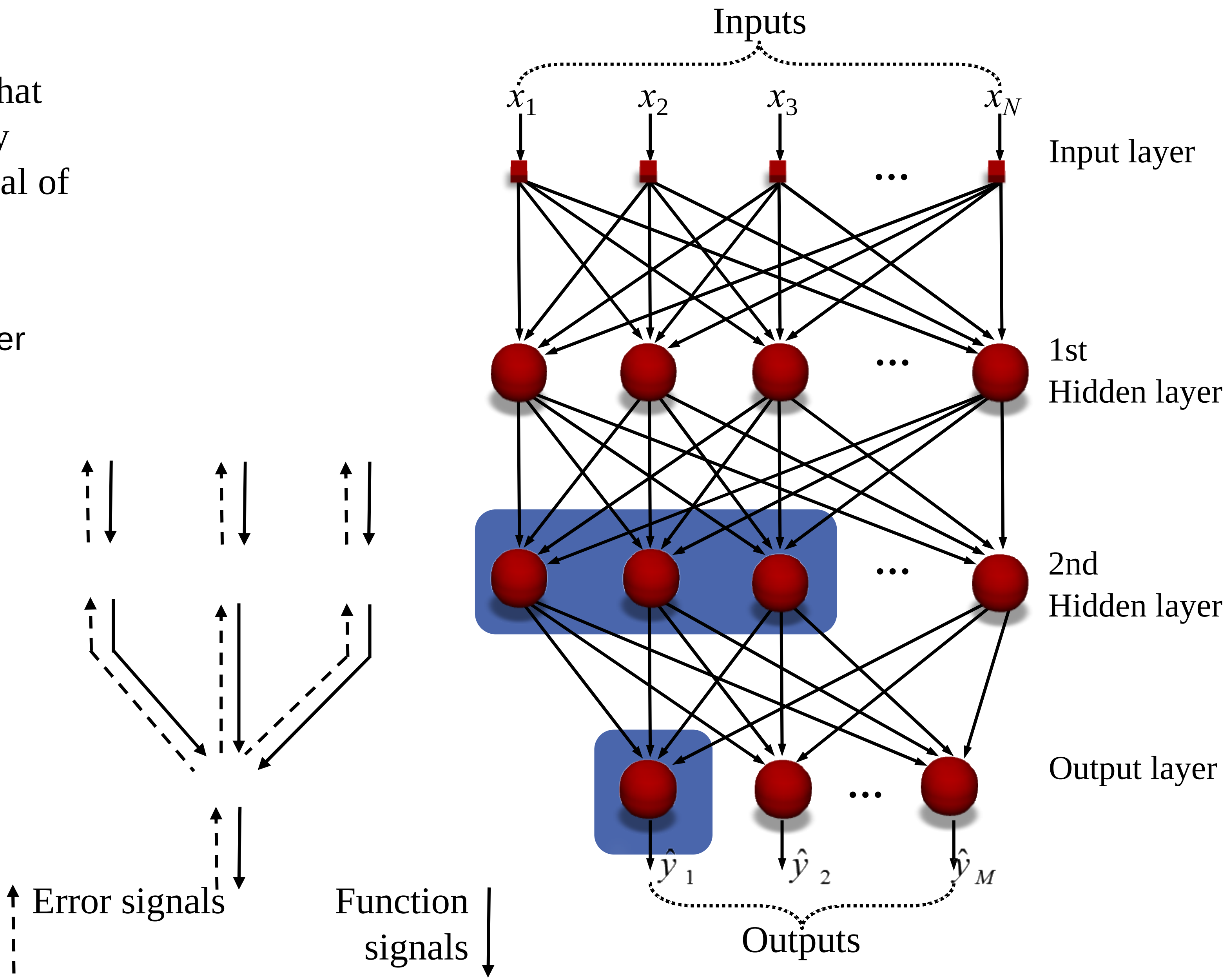
- **Learning or training** is a fundamental **capability** of NNs, which allows them to **learn from their environment and improve their behaviour**.

- The neural network **learns by adjusting its weights and bias (threshold)** iteratively to yield the desired output.

Multi-Layer Perceptrons

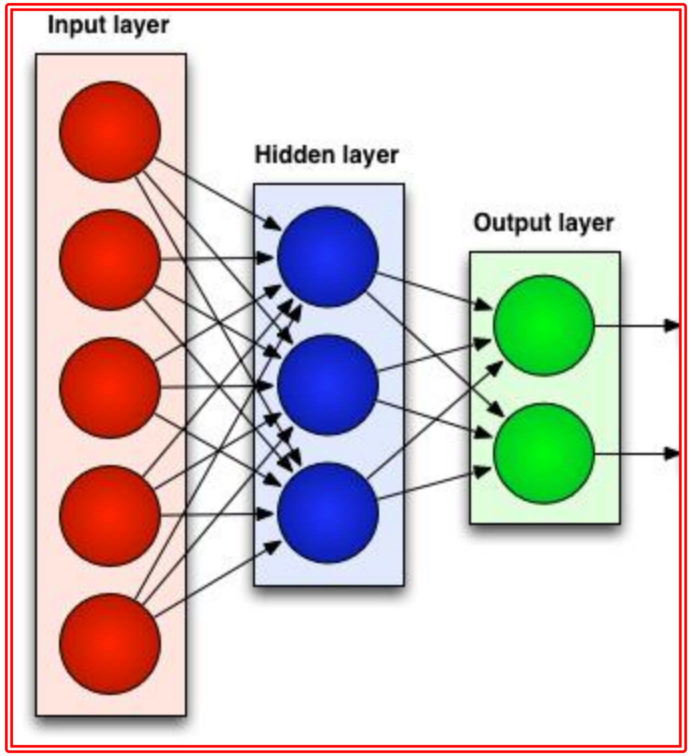
Backpropagation is the learning mechanism that allows the Multilayer Perceptron to iteratively adjust the weights in the network, with the goal of minimizing the cost function.

- ❑ Feedforward step:
  - ➢ an input pattern is applied to the input layer and its effect propagates, layer by layer, through the network until an output is produced.
  - ➢ the network's actual output value is then compared to the expected output, and an error signal is computed for each of the output nodes.
- ❑ Backward step:
  - ➢ the output error signals are transmitted backwards from the output layer to each node in the hidden layer that immediately contributed to the output layer.

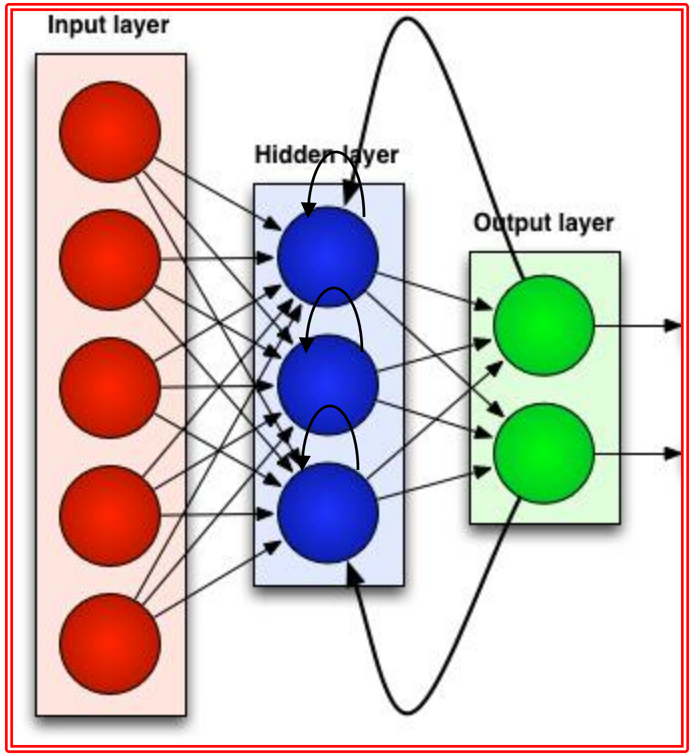


Neural Networks: Static/Dynamic

- Static networks: Feedforward neural networks
  - They learn a static I/O mapping, $\boldsymbol{Y}=\boldsymbol{f}(\boldsymbol{X})$, $\boldsymbol{X}$ and $\boldsymbol{Y}$ static patterns (arrays)

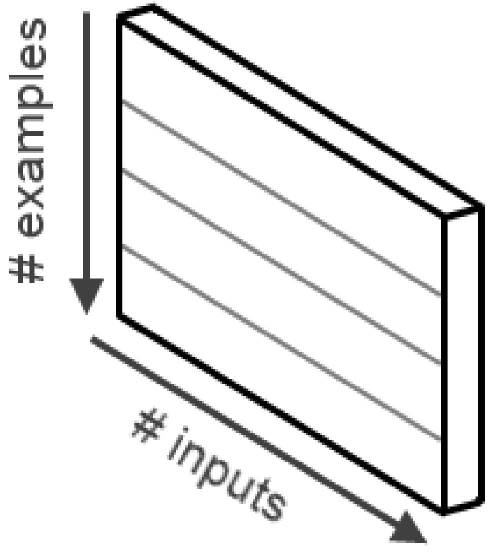


- Dynamic networks: Recurrent neural networks
  - They learn a dynamic I/O mapping, $\boldsymbol{Y}(t)=\boldsymbol{f}(t,\boldsymbol{X}(t))$, $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are time–varying patterns



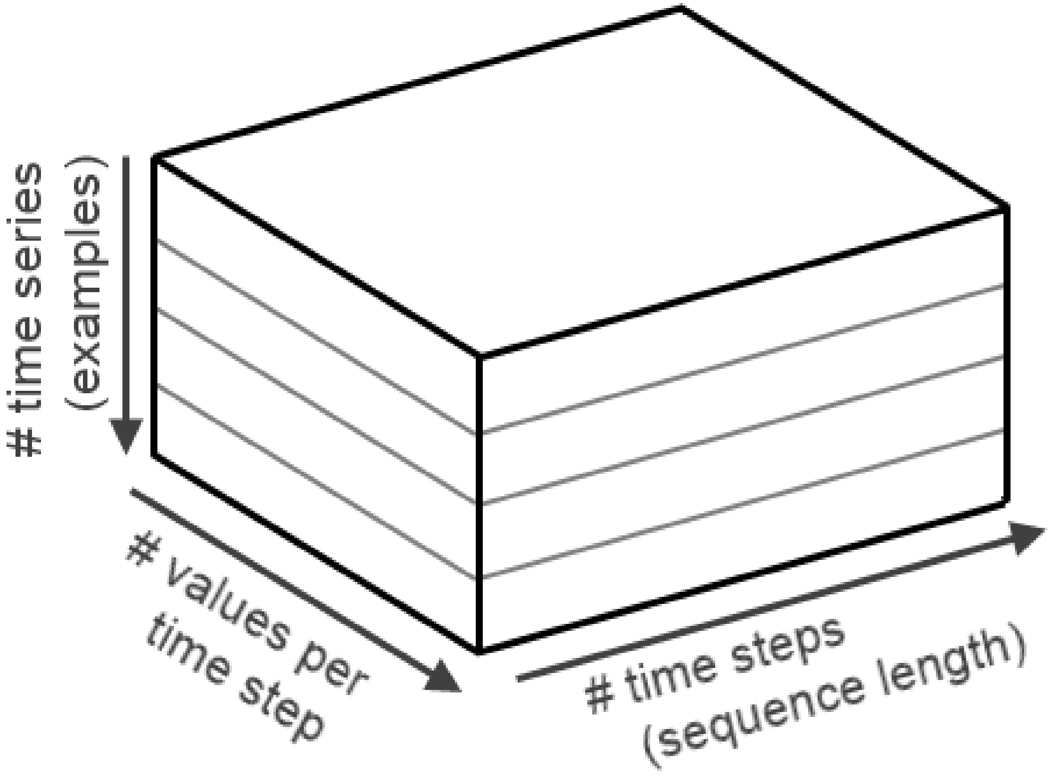
Static Feedforward Networks vs Recurrent Networks

Static patterns



[examples, features]

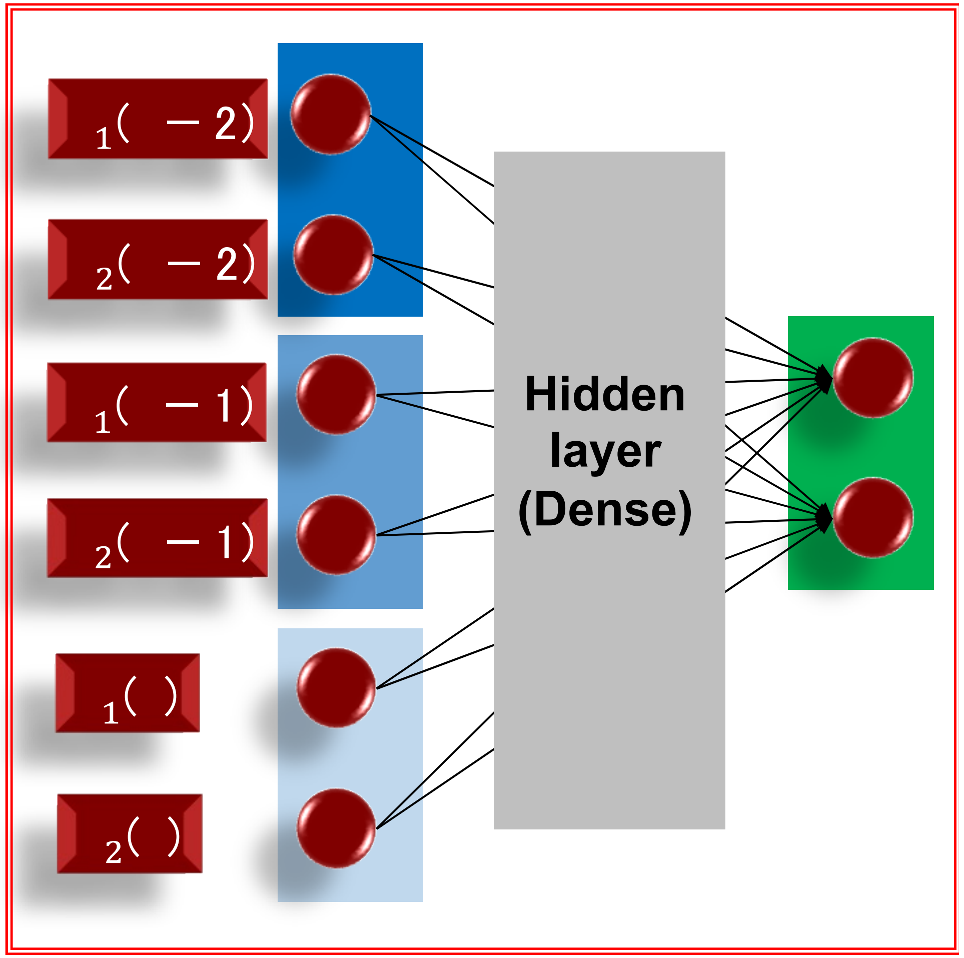
Time-varying patterns



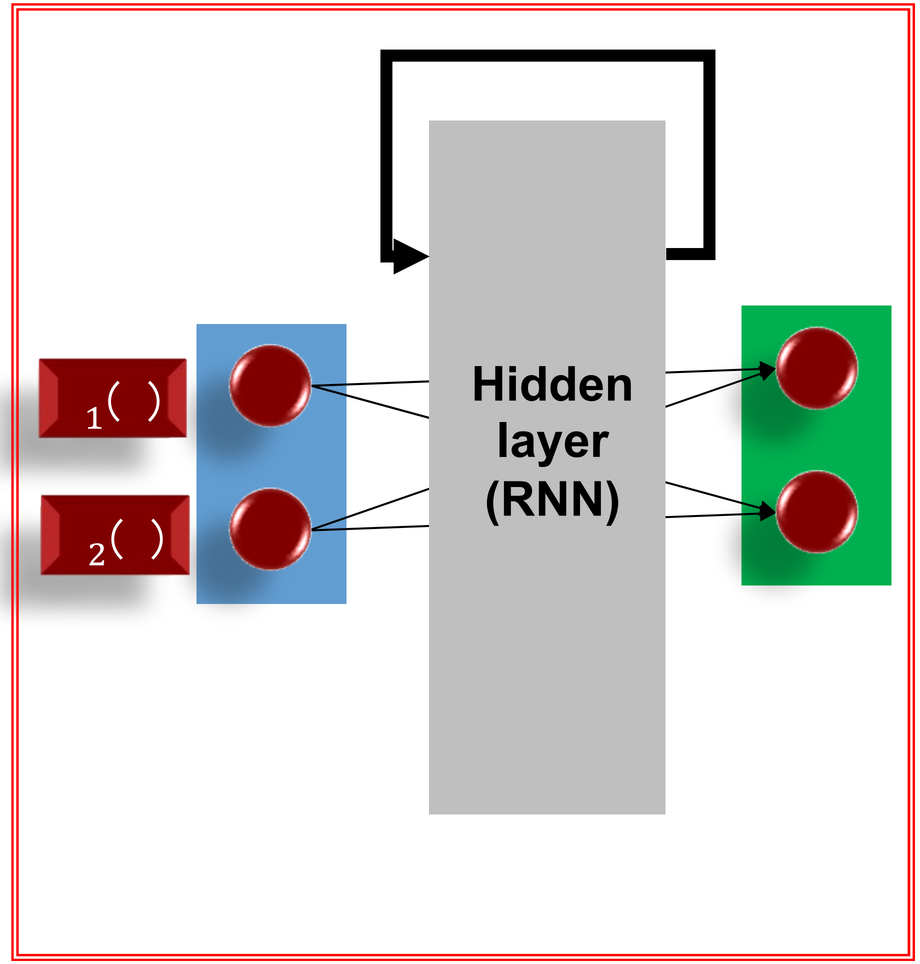
[examples, timesteps, features]

Neural Networks for timeseries

- Static networks: Feedforward neural networks



- Dynamic networks: Recurrent neural networks



Real World Problems - Applications

Reading List

- Chondrodima E., et al. (2023) **An Efficient LSTM Neural Network-Based Framework for Vessel Location Forecasting**. IEEE Transactions on Intelligent Transportation Systems
- Mandalis P., et al. (2023) **Towards a Unified Vessel Traffic Flow Forecasting Framework**. Proc. IEEE Int. Workshop BMDA.
- Chondrodima E., et al. (2022) **Machine Learning Models for Vessel Route Forecasting: An Experimental Comparison**. Proc. 23rd IEEE Int. Conf. MDM.
- Mandalis P., et al. (2022) **Machine Learning Models for Vessel Traffic Flow Forecasting: An Experimental Comparison**. Proc. 3rd IEEE Int. Workshop MBDW.
- Tritsarolis A., et al. (2022) **Vessel Collision Risk Assessment using AIS Data: A Machine Learning Approach**. Proc. 3rd IEEE Int. Workshop MBDW.
- Tampakis P., et al. (2022) **i4sea: a big data platform for sea area monitoring and analysis of fishing vessels activity**. Geo-Spatial Information Science.
- Tampakis P., et al. (2022) **Sea area monitoring and analysis of fishing vessels activity: The i4sea big data platform**. Proc. 21st IEEE Int. Conf. MDM.
- Troupiotis-Kapeliaris A., et al. (2022) **Data Driven Digital Twins for the Maritime Domain**. Proc. 21st IEEE Int. Conf. MDM.
- Patroumpas K., et al. (2020) **Trajectory Detection and Summarization over Surveillance Data Streams**. Big Data Analytics for Time-Critical Mobility Forecasting
- Petrou P., et al. (2019) **ARGO: A Big Data Framework for Online Trajectory Prediction**. Proc. 16th Int. Conf. SSTD.

Open-source: github.com/DataStories-UniPi

Real World Problems – Applications:
Vessel Route Forecasting
Fishing Vessels Activity Prediction

Vessel Route Forecasting - Motivation

Vast spread of AIS-enabled maritime fleet & Maritime Transport Systems (MTS) → Motivation for several analytics (incl. **forecasting)** tasks

Accurate and timely **Vessel Route Forecasting** (**VRF**):

- is critical for safety at sea
- can assist shipping industry in improving travel efficiency
- has a wide range of applications, such as accurate ETA calculation, collision / traffic jam assessment, etc.
- is challenging due to complex and dynamic maritime traffic conditions

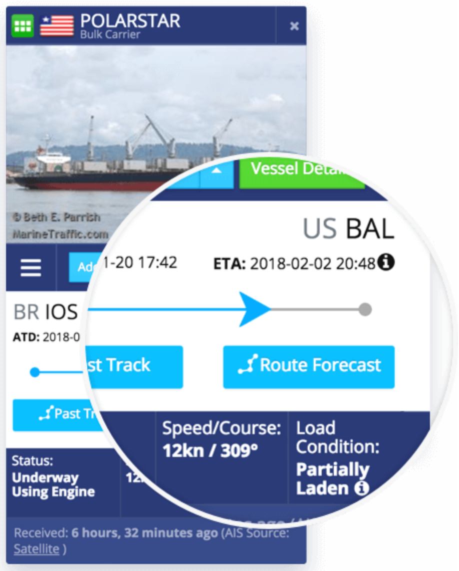


image source: marinetraffic.com

Our Contribution vs. Related Work

Various methods have been proposed to address VRF, e.g. [1-3]. However:

- Due to limited comparison analysis, it is hard to evaluate their robustness for the purpose of MTS operational usage
- Using preprocessing (e.g., interpolation) to create points at a fixed sampling rate can lead to a) higher computational load, and b) poor model predictions [4]

Our work:

- ✓ **Examines the most popular ML methods** to address the VRF problem & to **provide a fair comparison study**.
- ✓ **Examines the effect of different sea areas**, through an experimental setup that includes 3 real-world maritime datasets.
- ✓ **Enhances ML method's prediction accuracy** though the Trajectory Data Augmentation (TDA) method tailored to trajectory learning.
- ✓ **Addresses the sparsity and variable sampling rate of vessel data** through the spatiotemporal-aware processing mechanism.

[1] Valsamis et.al. (2017) Employing traditional machine learning algorithms for big data streams analysis: The case of object trajectory prediction. J. Syst. Softw.
[2] Tu et.al. (2018) Exploiting ais data for intelligent maritime navigation: A comprehensive survey from data to methodology. IEEE Trans. Intell. Transp. Syst.
[3] Wang et.al. (2020) Trajectory forecasting with neural networks: An empirical evaluation and a new hybrid model. IEEE Trans. Intell. Transp. Syst.
[4] Weerakody et.al. (2021) A review of irregular time series data handling with gated recurrent neural networks. Neurocomputing.

Problem Formulation - Vessel Route Forecasting

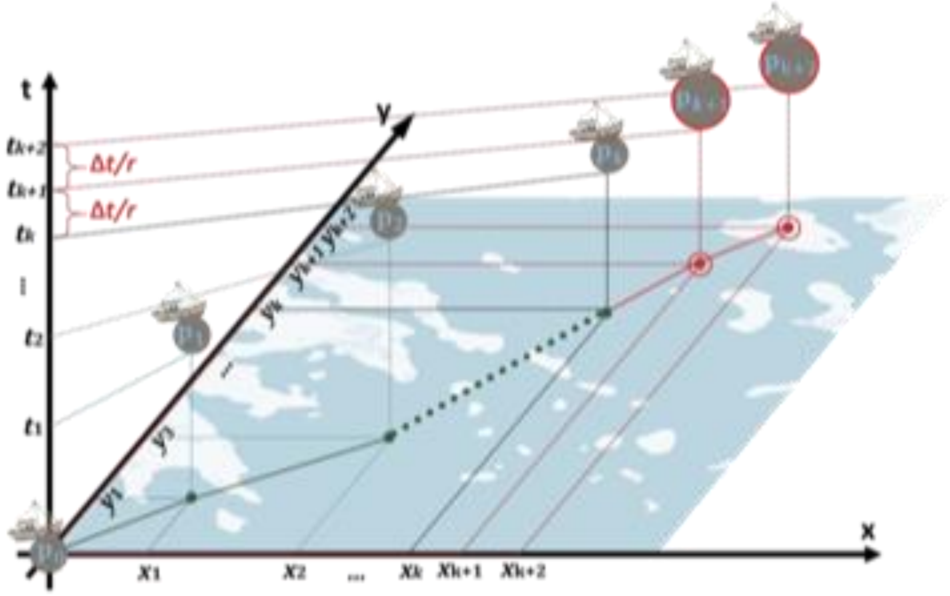
The **Vessel Route Forecasting (VRF) problem** over a dataset composed of vessel trajectories

Given:

- a vessel's trajectory $[(\mathbf{p}_0,\mathbf{t}_0), \ldots, (\mathbf{p}_k, \mathbf{t}_k)]$ consisting of ***k*** transitions at (irregular) timepoints,
- a time duration (prediction horizon) ***Δt***
- a number of transitions ***r***

Predict:

- the **vessel's future trajectory** $[(\mathbf{p}_{k+1},\mathbf{t}_{k+1}), \ldots, (\mathbf{p}_{k+r}, \mathbf{t}_{k+r})]$ consisting of ***r*** transitions at (fixed) timepoints, i.e., with sampling rate equal to ***Δt/r***



* Note: ***r***=1 // Vessel Location Forecasting (VLF)

Problem Formulation - Activity Prediction

The **Activity Prediction (AP) problem** of fishing vessels:

Given:

- the **vessel's future trajectory (**predicted using VRF method)

Predict:

- the **vessel's future activity**, either (a) at timestamp $\mathbf{t_i} + \mathbf{\Delta t}$ or (b) until timestamp $\mathbf{t_i} + \mathbf{\Delta t}$, where **activity** is one of {Mooring, Fishing, Steaming}

- Tampakis P., et al. (2022) **i4sea: a big data platform for sea area monitoring and analysis of fishing vessels activity**. Geo-Spatial Information Science.
- Tampakis P., et al. (2022) **Sea area monitoring and analysis of fishing vessels activity: The i4sea big data platform**. Proc. 21st IEEE Int. Conf. MDM.

Activity Prediction Definitions/Rules

Mooring:
Vessels are within/close to ports.

Fishing:
Vessels sailing away from ports with lower speed.

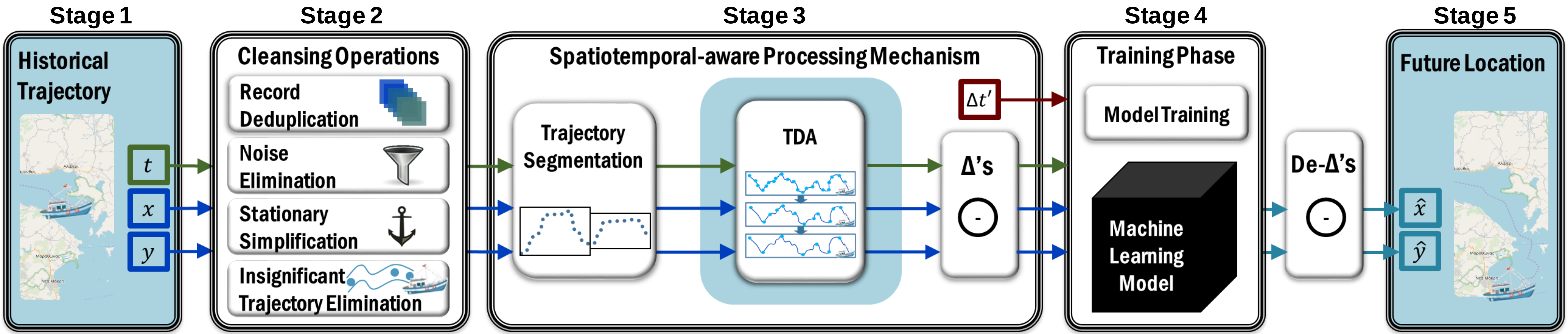
Steaming:
Vessels sailing away from ports with higher speed.

Detailed AP rules

Ship activity	Ship type	Distance from port	Speed	Period
Mooring	Trawlers	< 3 n.m.	-	-
	Purse seiners	< 3 n.m.	-	-
Fishing	Trawlers	> 3 n.m.	< 4knots	-
	Purse seiners	> 3 n.m.	< 1knots	Month: Apr.-Oct., Hour (UTC): [17:00 - 24:00] & [00:00 - 01:00] Month: Nov.-Dec. & Jan.-Mar., Hour (UTC): [17:00 - 24:00] & [00:00 - 02:00]
Steaming	Trawlers	> 3 n.m.	> 4knots	-
	Purse seiners	> 3 n.m.	> 1knots	-
			< 1knots	Month: Apr.-Oct., Hour (UTC): [01:00, 17:00] Month: Nov.-Dec. & Jan.-Mar., Hour (UTC): [02:00, 17:00]

Vessel Location/Route Forecasting Framework

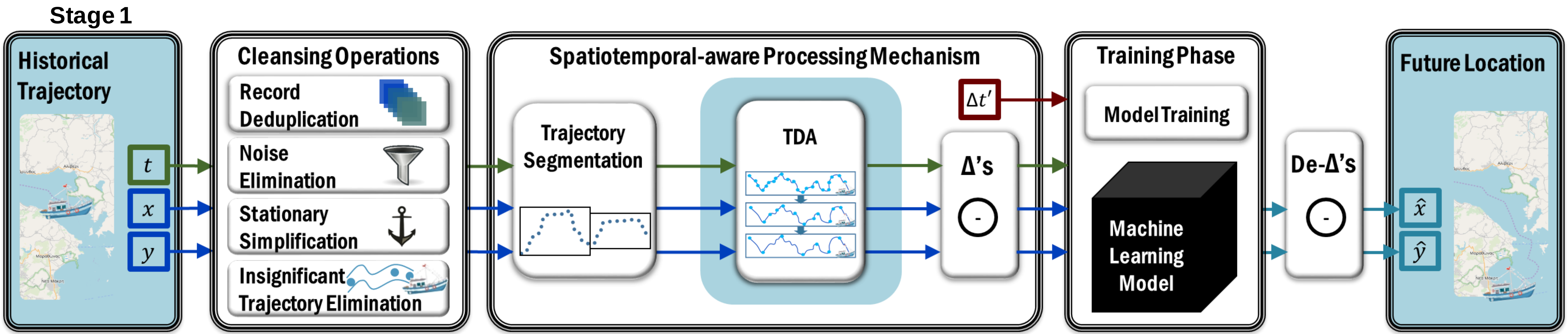
Overview of the proposed framework



- Open Source: https://github.com/DataStories-UniPi/VLF_VRF
- Chondrodima E., et al. (2023) **An Efficient LSTM Neural Network-Based Framework for Vessel Location Forecasting**. IEEE Transactions on Intelligent Transportation Systems
- Chondrodima E., et al. (2022) **Machine Learning Models for Vessel Route Forecasting: An Experimental Comparison**. Proc. 23rd IEEE Int. Conf. MDM.

Vessel Location/Route Forecasting Framework

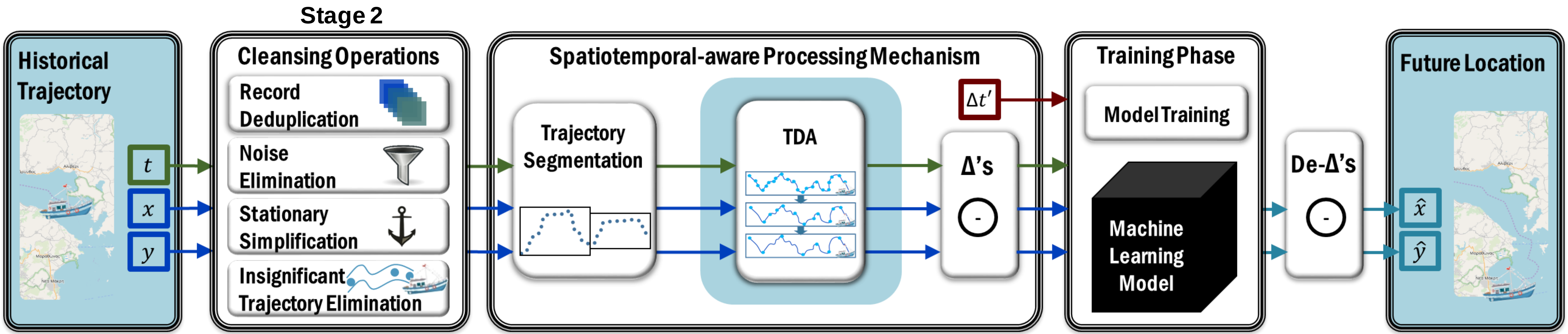
Overview of the proposed framework, consisting of five stages



… feeds the framework with vessels' positioning data

Vessel Location/Route Forecasting Framework

Overview of the proposed framework, consisting of five stages

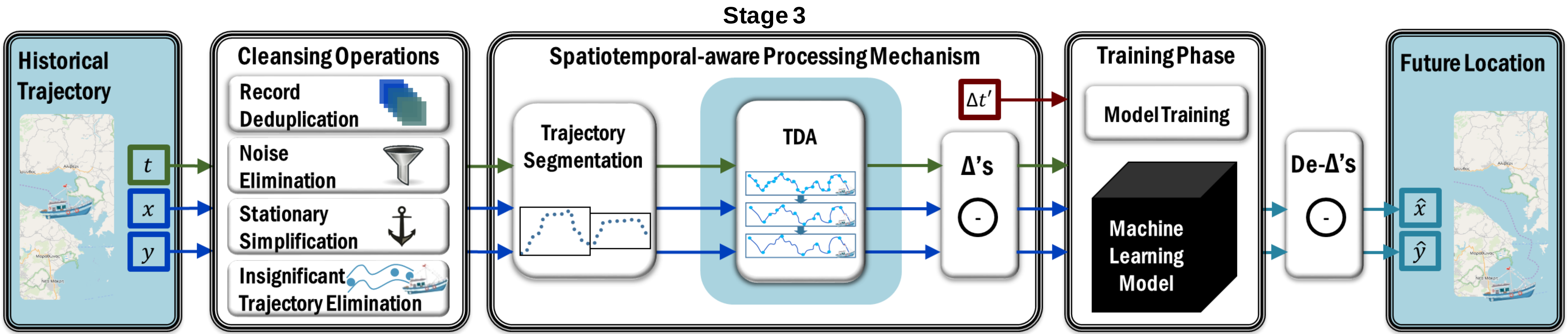


… performs data cleansing:

- Record deduplication: remove data records at the same timestamp or at timestamps differing less than 1 sec.
- Noise elimination: remove records corresponding to invalid speed (above 50 knots)
- Stationery simplification: remove records corresponding to speed that indicates immobility, (below 0.1 knots)
- Insignificant trajectory elimination: eliminate trajectories with low number of points (less than 20 points)

Vessel Location/Route Forecasting Framework

Overview of the proposed framework, consisting of five stages



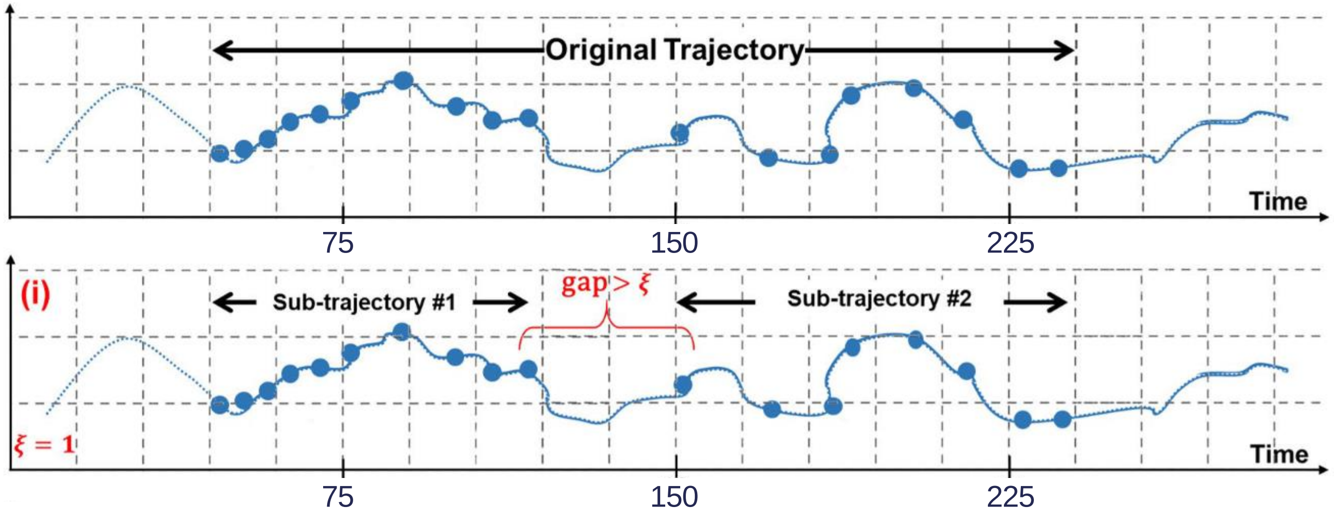
… includes the **spatiotemporal-aware processing mechanism**, whose purpose two-fold:

a) segments sparse trajectories to non-sparse

Vessel Location/Route Forecasting Framework: Spatiotemporal-aware processing mechanism – Trajectory Segmentation Process

Trajectories are partitioned into sub-trajectories when :

i. the time interval between two consecutive vessel points exceeds **30 minutes**

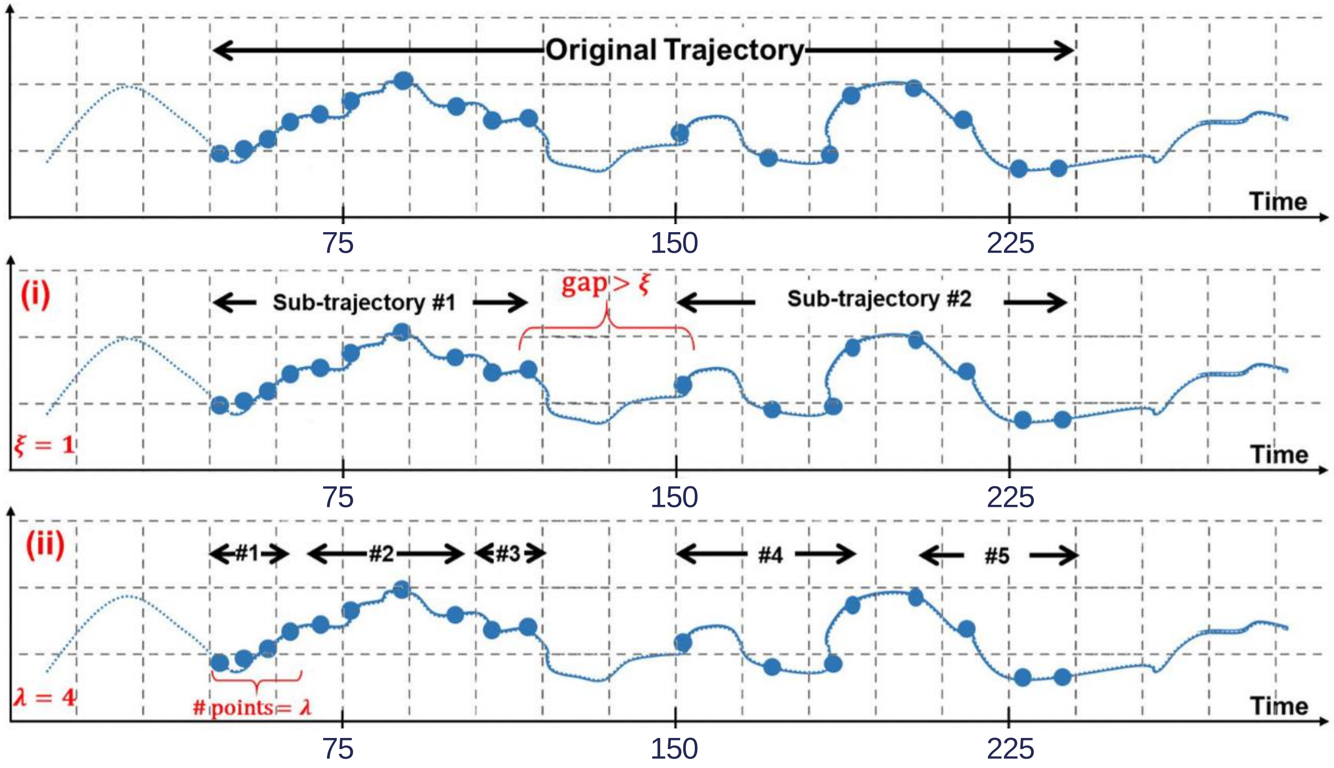


Vessel Location/Route Forecasting Framework: Spatiotemporal-aware processing mechanism – Trajectory Segmentation Process

Trajectories are partitioned into sub-trajectories when :

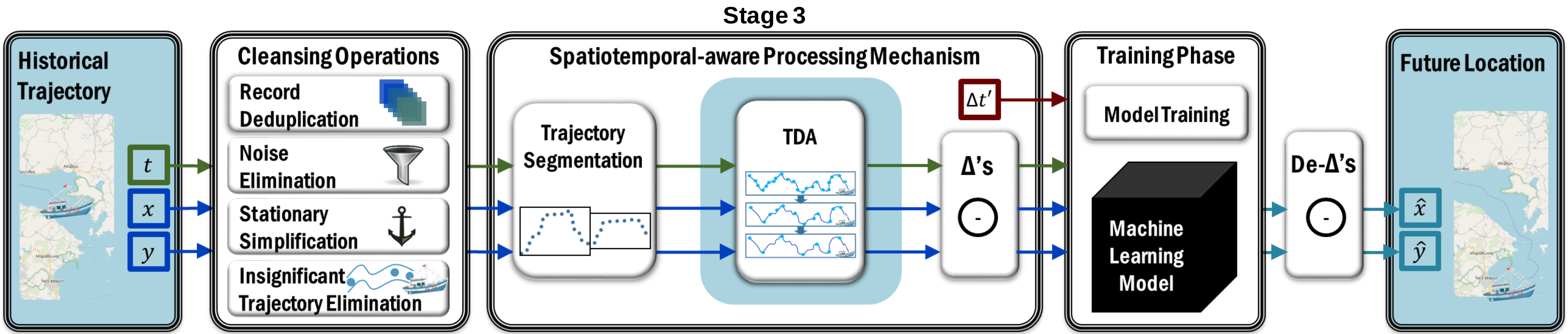
i. the time interval between two consecutive vessel points exceeds **30 minutes**

ii. the length of the resulting trajectory exceeds **1000 points**



Vessel Location/Route Forecasting Framework

Overview of the proposed framework, consisting of five stages



… includes the spatiotemporal-aware processing mechanism, whose purpose two-fold:

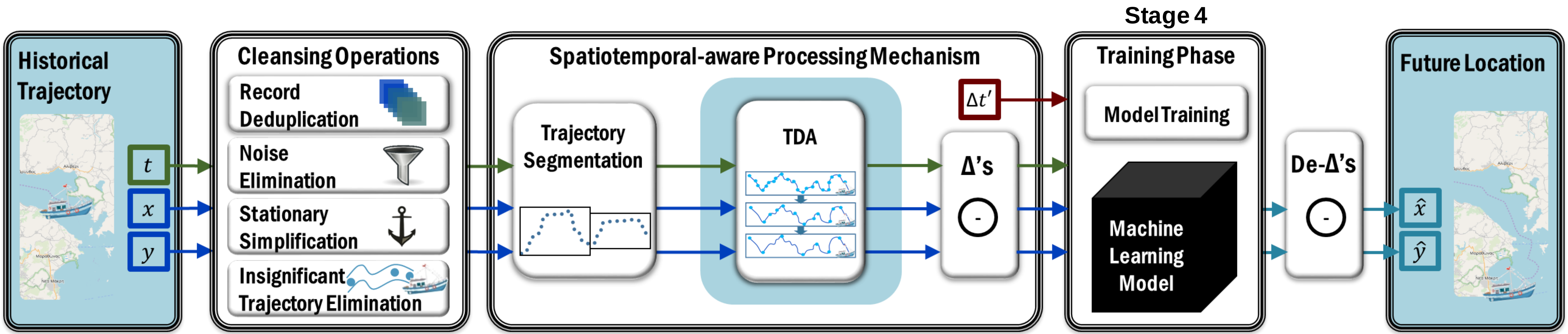
a) segments sparse trajectories to non-sparse

b) transforms asynchronous time sampled spatiotemporal information to a **representation suitable for RNN** models by:

- using **Trajectory Data Augmentation (TDA)**, which exploits on **Douglas-Peucker simplification algorithm**
- **converting** the time and the spatial **information** of each vessel **to differences**

Vessel Location/Route Forecasting Framework

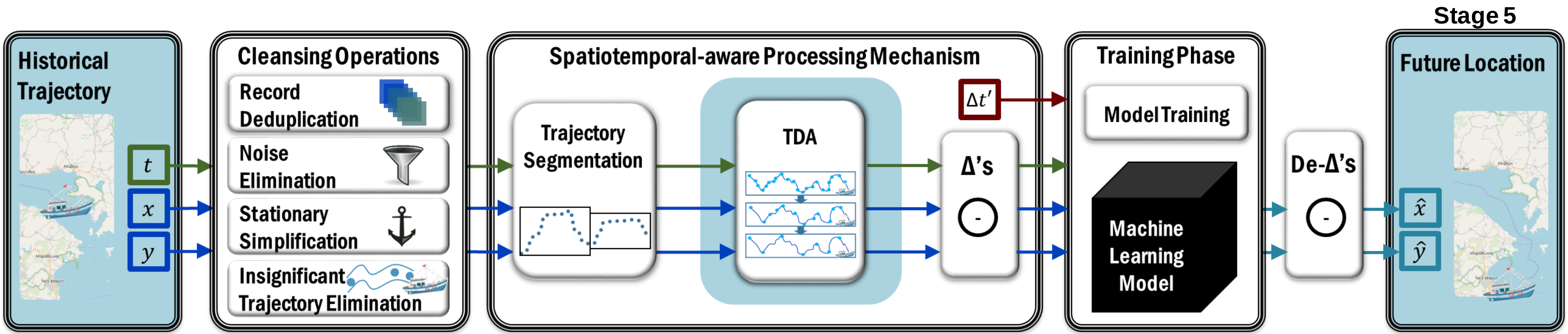
Overview of the proposed framework, consisting of five stages



… trains the model by using the desired time horizon and provides predictions

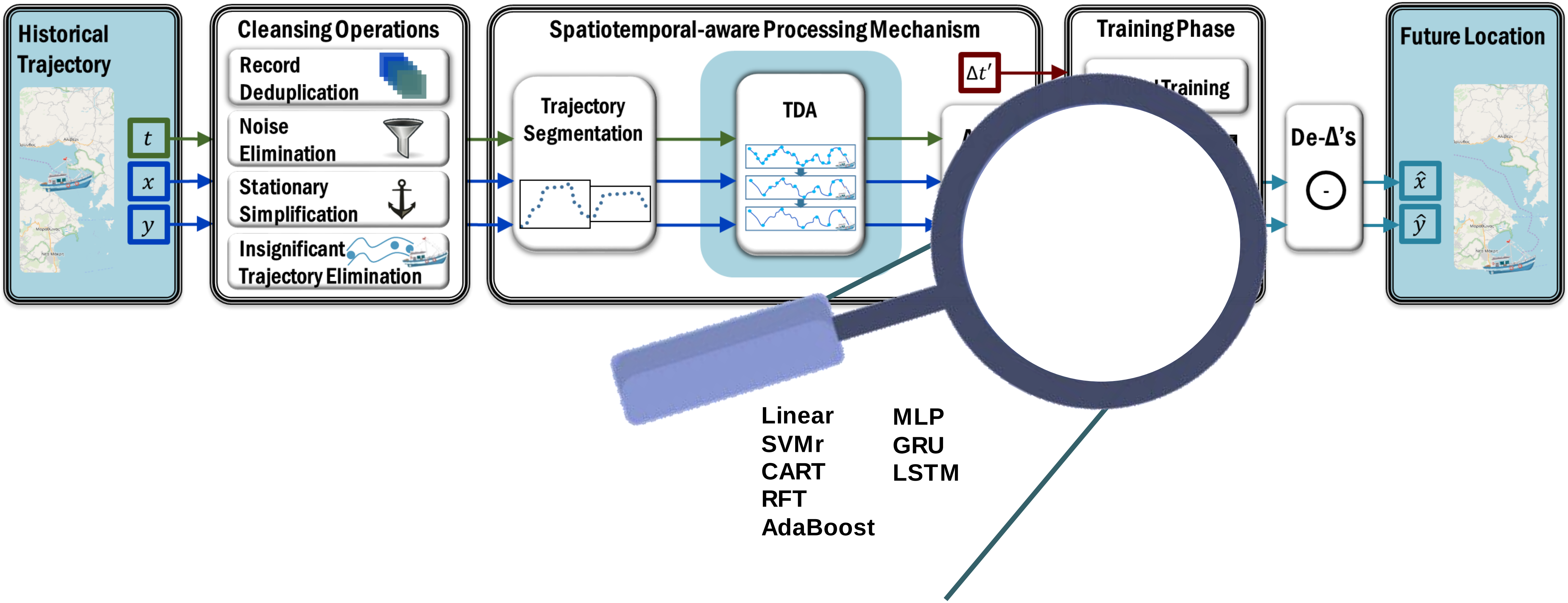
Vessel Location/Route Forecasting Framework

Overview of the proposed framework, consisting of five stages



… transforms the model's output and provides the predicted coordinates

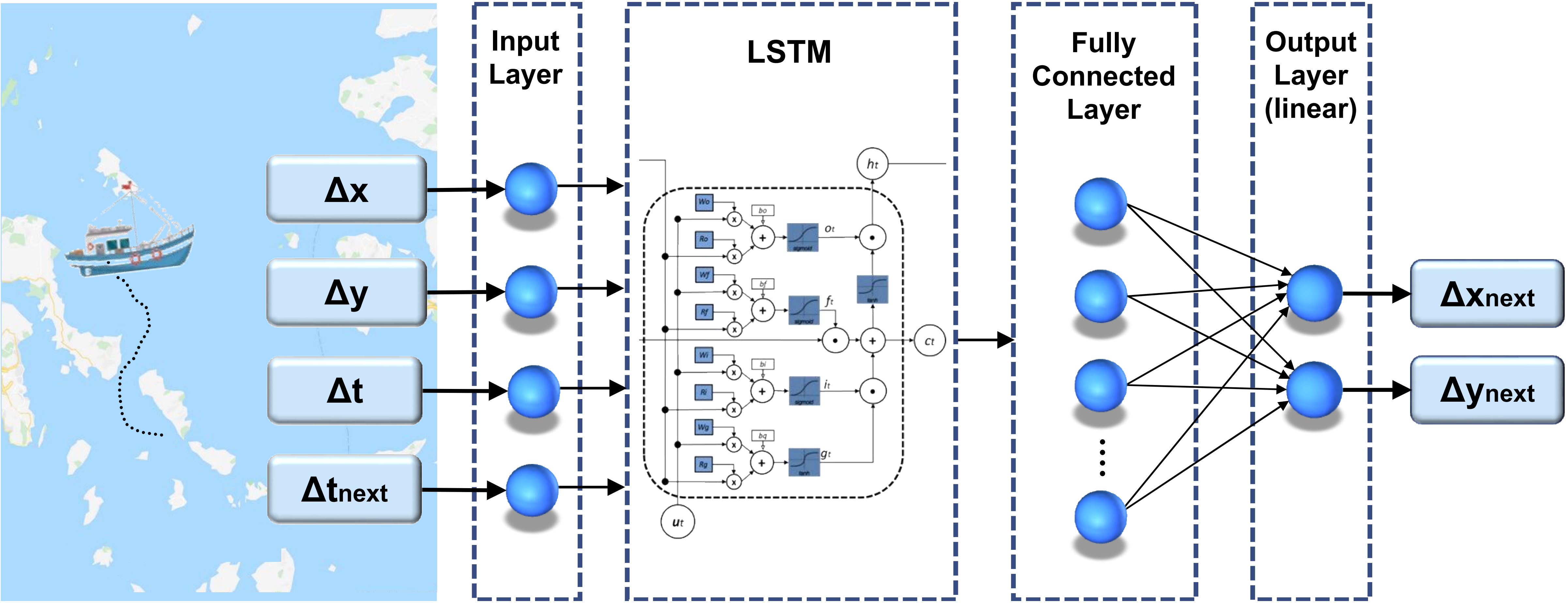
Vessel Location/Route Forecasting Framework



Linear
SVMr
CART
RFT
AdaBoost

MLP
GRU
LSTM

Vessel Location/Route Forecasting Framework: NN architecture

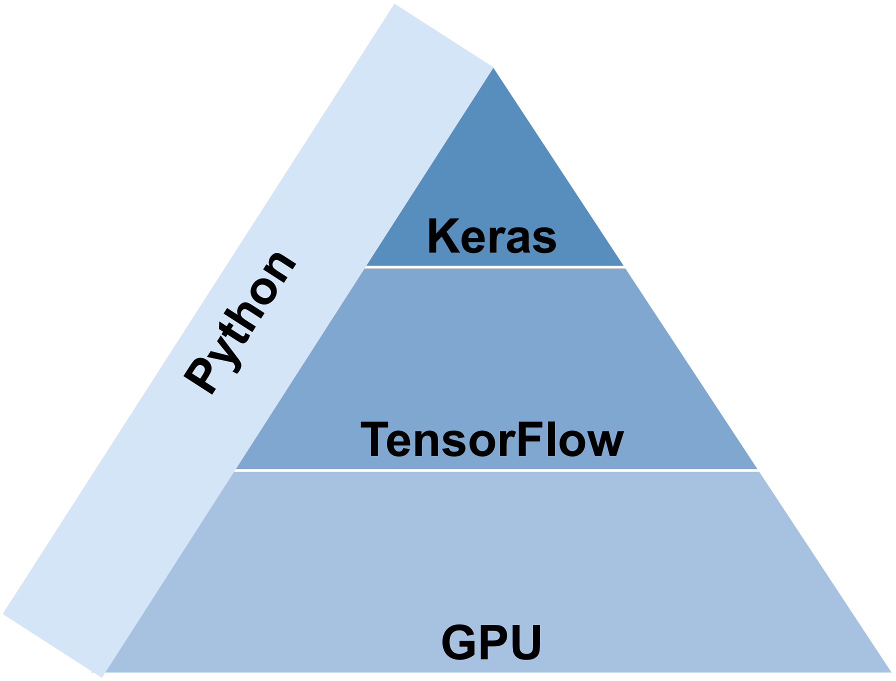


i4sea platform - Prediction Module: Tool Implementation

The NN model learns in an offline mode and predicts vessels locations in an online-streaming mode by applying the trained model

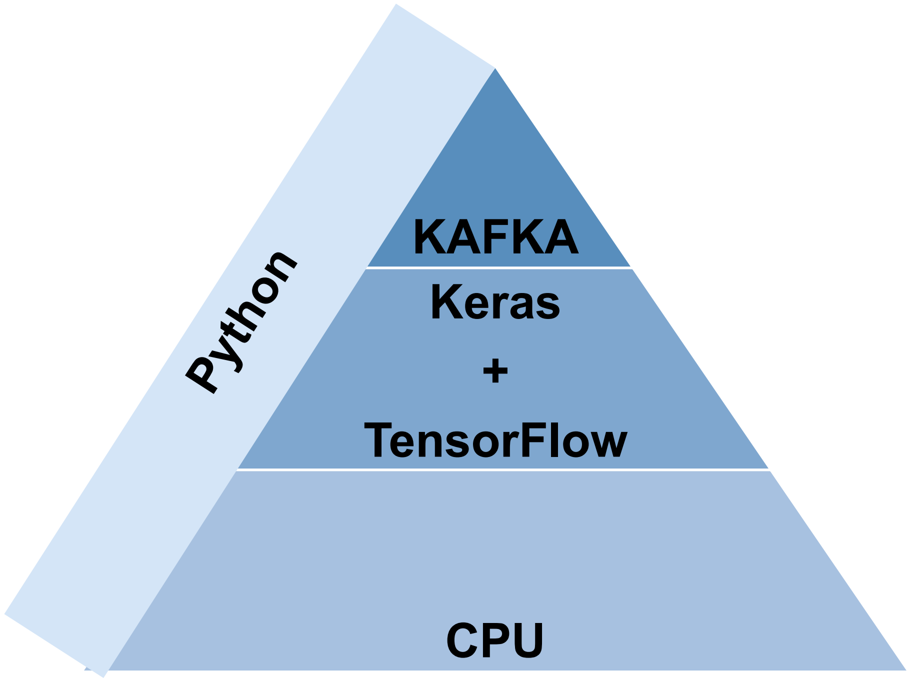
Training

(offline)



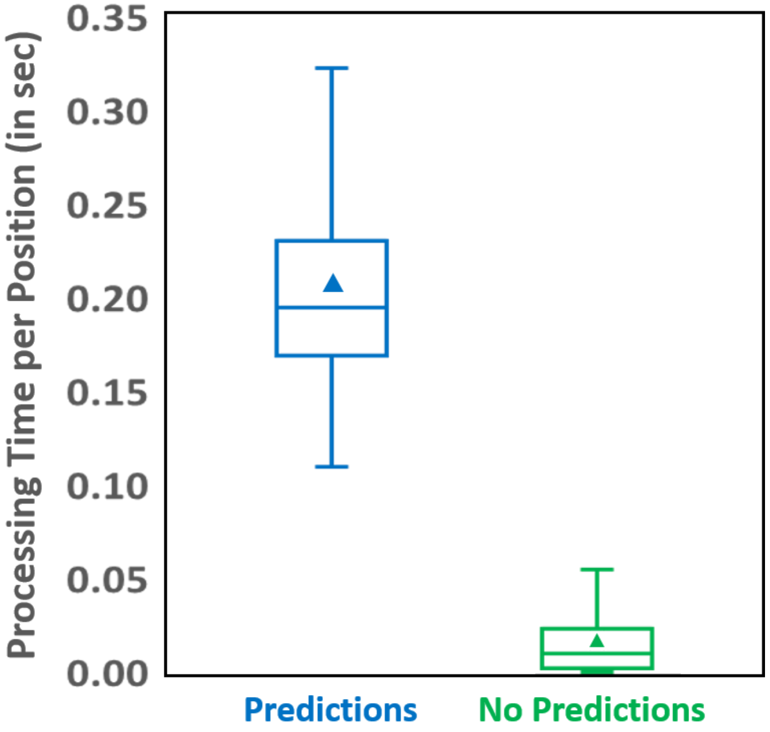
Prediction

(online-streaming)



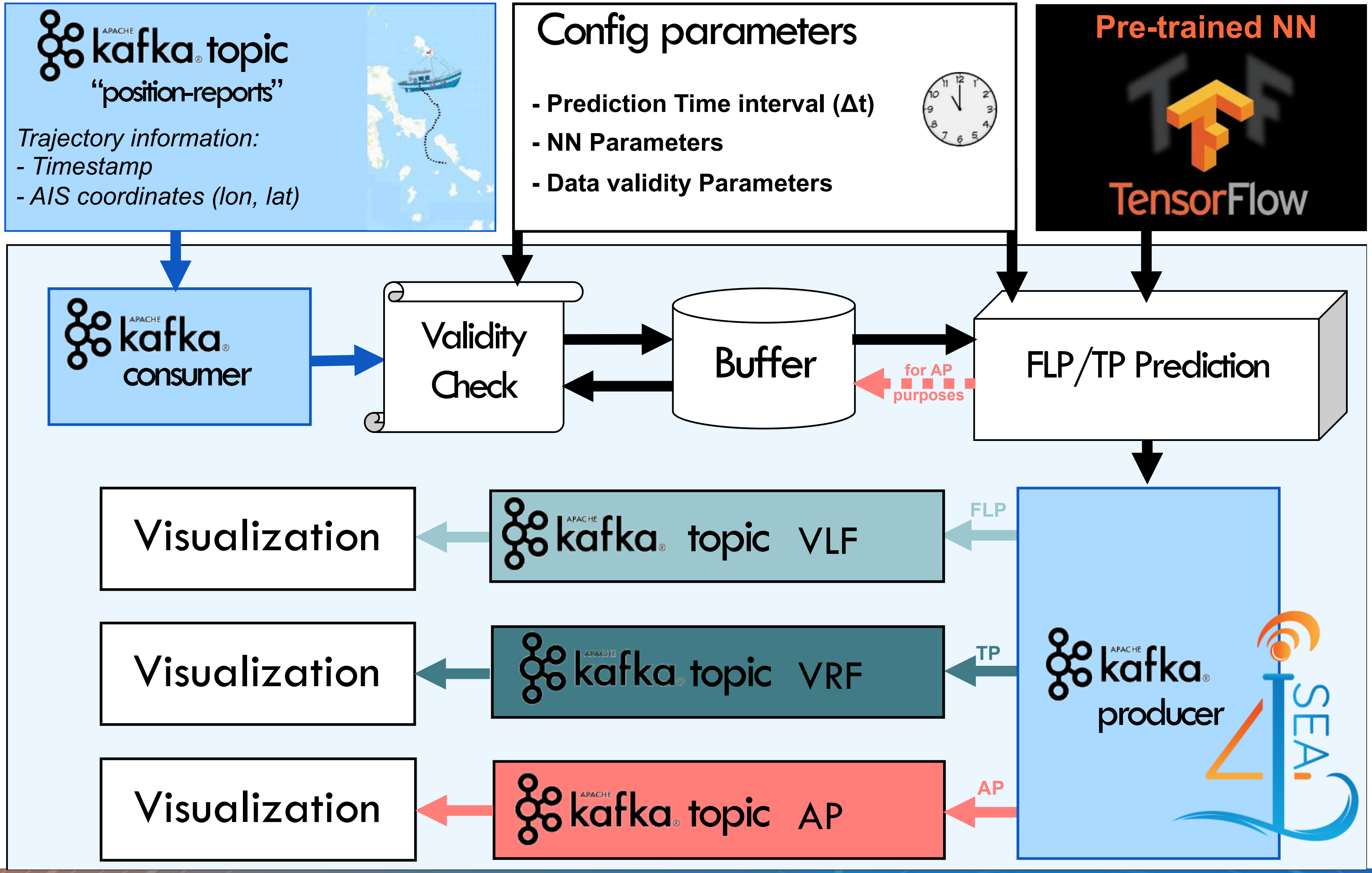
i4sea platform - Prediction Module: Online Performance

	Runtime (sec) for 1 prediction
Min	0.1559
Max	0.4720
Median	0.2377
Mean	0.2387

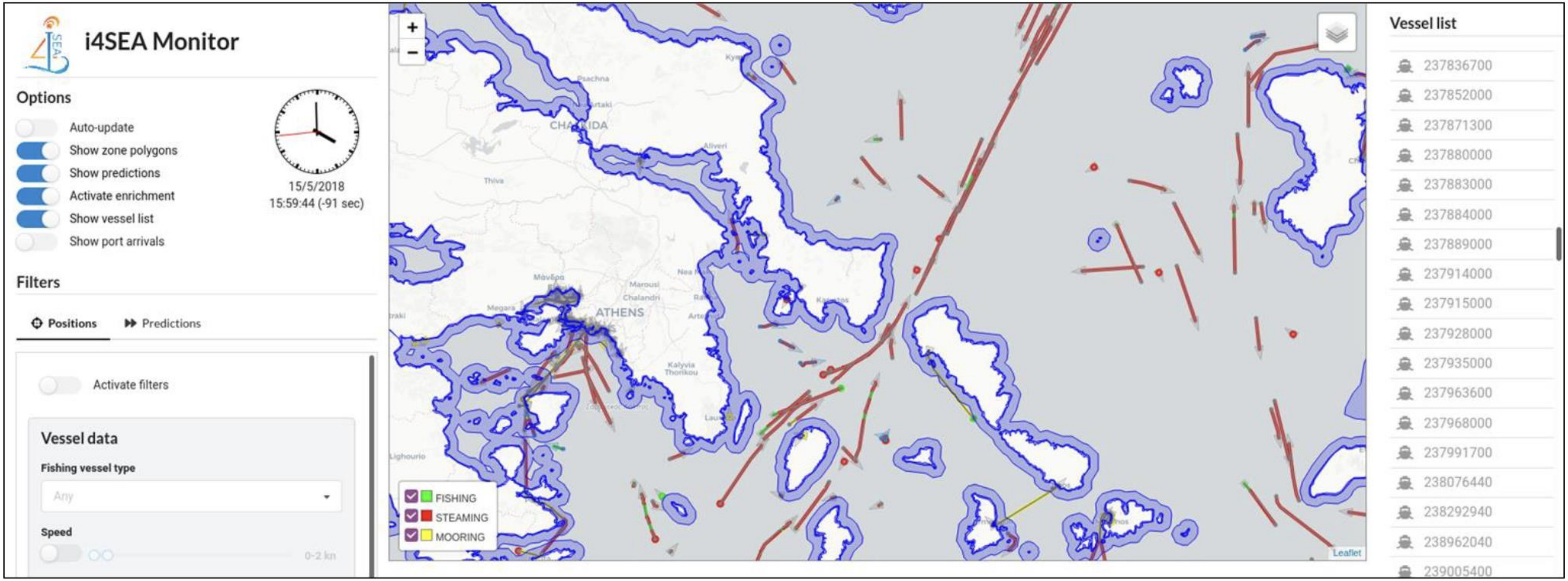


i4sea platform - Prediction Module: Online Process with pre-trained NN model

Overview of i4sea platform - Prediction Module workflow



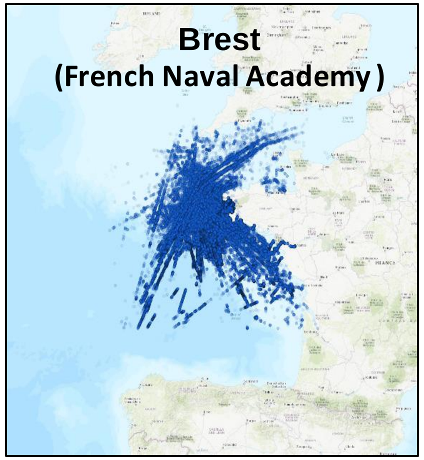
i4sea Platform - Monitor

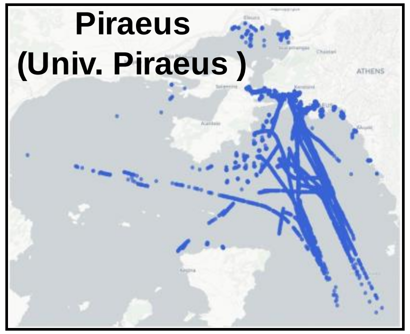


This demo shows future location predictions of three different vessels in a prediction time span of thirty minutes in fast forward in the i4sea platform.

Experimental Setup

- For our experimentation, we used real-world AIS datasets
- Experimental protocol:
  - training (50%), validation (25%), and testing (25%) randomly allocated
  - model parameters optimized through intermediate experiments





Dataset	U.S. West Coast	Brest	Aegean-Sea	Aegean-Cyclades	Piraeus
Provider	MarineCadastre	French Naval Academy	MarineTraffic	MarineTraffic	Univ.Piraeus
Time frame	1 month (01–30/11/2018)	6 months (01/10/2015-31/03/2016)	1 month (01–30/11/2018)	1 month (01–30/11/2018)	1 day (3/7/2018)
# of records	10,671,963	16,311,185	1,289,642	1,720,368	455,145
# of distinct vessels	1122	5041	2854	2645	361
Sampling rate (avg.)	< 1 min	< 1 min	~ 2.5 min	~ 2.5 min	~ 5 min

Experimental Results: Vessel Location Forecasting

- Results for the implemented methods …
  - were evaluated in the testing set, in terms of Euclidean distance between the original and the predicted points
  - include the best result, followed by the average and standard deviation values from the 10 runs in parentheses.

PREDICTION RESULTS - IMPLEMENTED METHODS AND RELATED WORK (UNIT: METERS)

Data	Method	Error Distance (meters) per Prediction interval (min.)			
Aegean Sea		4	10	20	30
	VLFF-LSTM	36 (36±1)	190 (207±10)	381 (458±46)	895 (1081±168)
	MLP+LSTM*	177 (189±7)	555 (585±26)	955 (1025±61)	1729 (1801±179)
	MLP[a]	153	652	983	1721
U.S. West Coast		15	30	45	60
	VLFF-LSTM	523 (539±20)	816 (941±107)	1900 (2179±241)	2617 (3554±998)
	MLP+LSTM*	1232 (1414±187)	2260 (2483±176)	3472 (3569±106)	4722 (4827±111)
	ELM[b]	1235	2789	4808	7201
Brest Area		4	8	16	32
	VLFF-LSTM	107 (113±4)	298 (311±16)	849 (860±17)	2400 (2454±65)
	MLP+LSTM*	901 (927±21)	1422 (1652±147)	2563 (2605±82)	4335 (4455±137)
	FLP-L[c]	1000	2000	5000	10000

[a] Valsamis et.al., "Employing traditional machine learning algorithms for bigdata streams analysis: The case of object trajectory prediction", Journal of Systems and Software, 2017.
[b] Tu et.al. "Exploiting ais data for intelligent maritime navigation: A comprehensivesurvey from data to methodology", IEEE Transactions on Intelligent Transportation Systems, 2018.
[c] Petrou et.al. "Online long-term trajectory prediction based on mined route patterns", International Workshop on Multiple-Aspect Analysis of Semantic Trajectories, 2019.
[*] Wang et.al., "Trajectory forecasting with neural networks: An empirical evaluation and a new hybrid model", IEEE Transactions on Intelligent Transportation Systems, 2020.

Experimental Results: Vessel Route Forecasting (VRF)

Quality measures:

- **Average displacement error (ADE)** – the average distance error for all predicted time steps
- **Final displacement error (FDE)** – the distance error at the final predicted time step

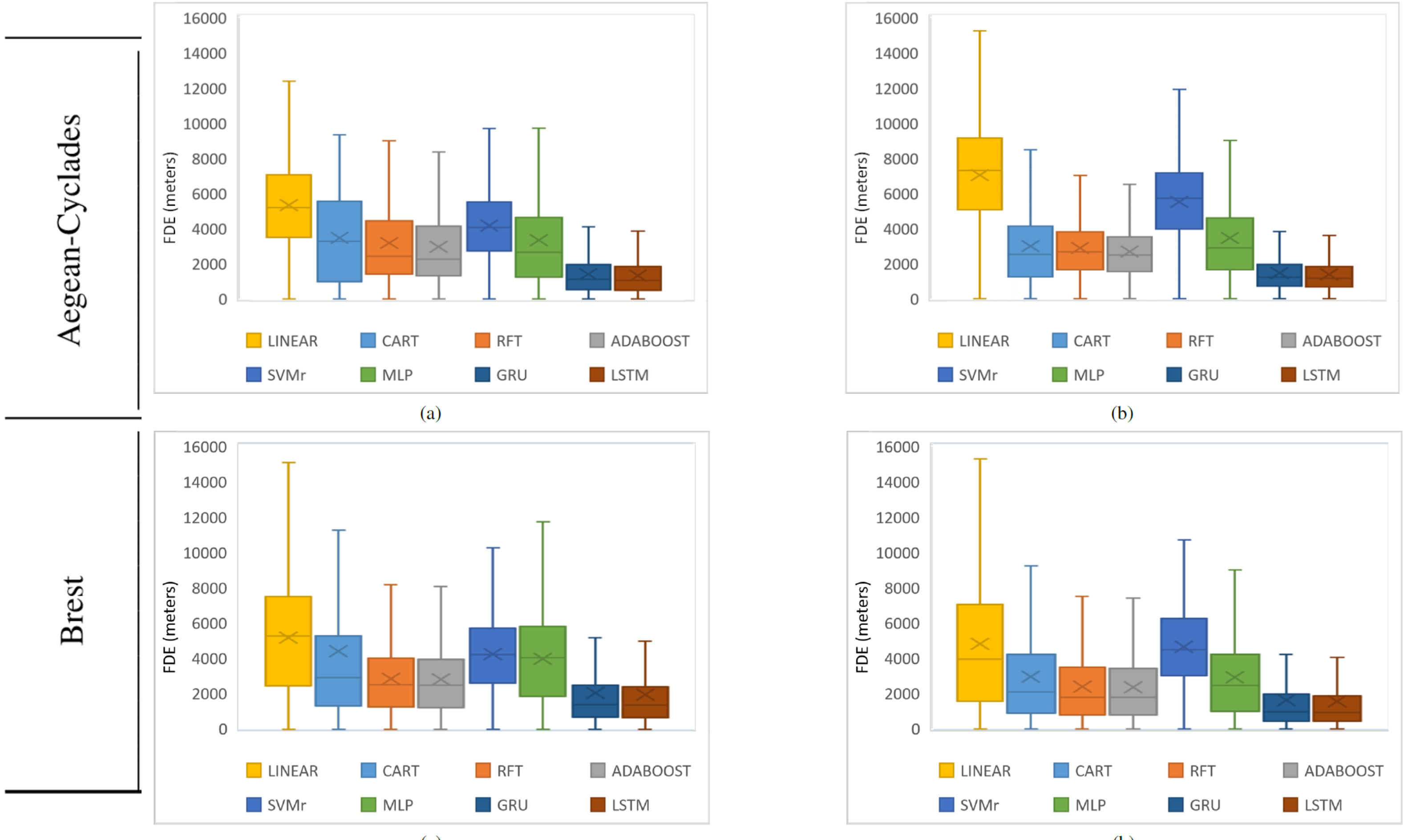
PREDICTION RESULTS FOR Δt UP TO 30 MIN. AND r UP TO 6 TRANSITIONS (UNIT: METERS)

Data	Method	ADE per Δt in min. for r=6						FDE(30 min)
		5	10	15	20	25	30	
Aegean-Cyclades	Linear	867	1717	2569	3420	4271	5121	9371
	CART	340	889	1481	1916	2335	2796	5102
	RFT	221	654	1114	1506	1911	2377	4709
	AdaBoost	230	640	984	1374	1785	2217	4376
	SVMr	638	1335	2223	2938	3706	4310	7328
	MLP	180	735	1290	1782	2264	2765	5270
	GRU	79	195	337	511	727	977	2229
	LSTM	**76**	**184**	**317**	**481**	**684**	**920**	**2097**
Brest	Linear	1158	1788	2412	3030	3642	4312	7666
	CART	571	1091	1679	2218	2708	3247	5945
	RFT	286	641	1016	1445	1852	2226	4094
	AdaBoost	252	610	983	1387	1782	2159	4041
	SVMr	697	1388	2008	2668	3276	3828	6591
	MLP	677	1067	1482	1936	2403	2894	5344
	GRU	241	466	710	959	1215	1485	2832
	LSTM	**239**	**440**	**663**	**899**	**1146**	**1408**	**2719**

Experimental Results: Vessel Route Forecasting (VRF) (cont.)

A closer look at FDE:

- distinct calculations regarding (a) eastings and (b) northings



(a)

(b)

(a)

(b)

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

DATA STORIES
Data Science Lab. @ University of Piraeus

Real World Problems – Applications: Vessel Traffic Flow Forecasting (VTFF)

Motivation

Vast spread of AIS-enabled maritime fleet & Maritime Transport Systems (MTS) → Motivation for several analytics (incl. **forecasting)** tasks

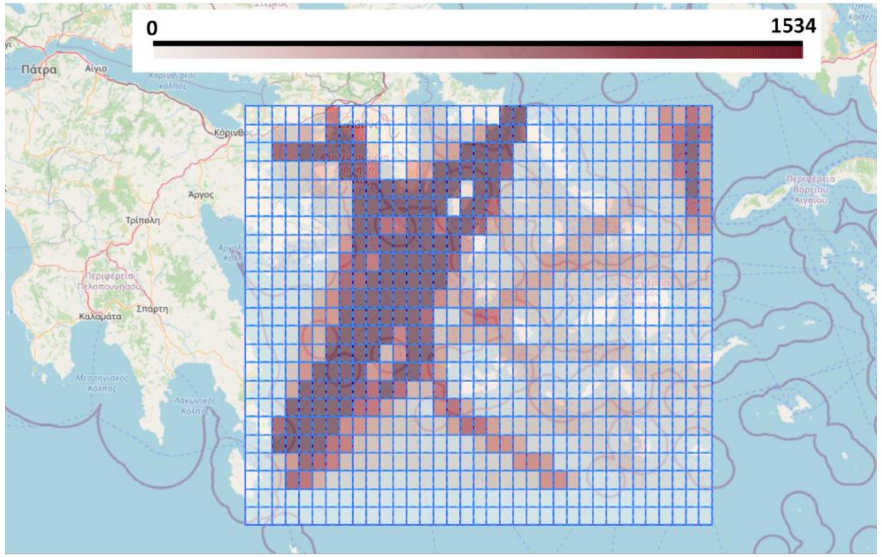


image source: [1]

Accurate **Vessel Traffic Flow Forecasting (VTFF)**:

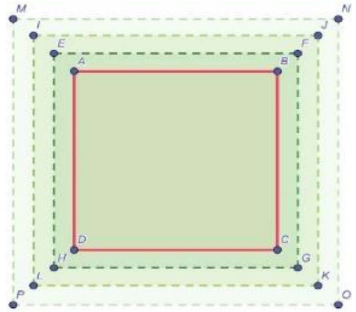
- is challenging due to the complex and dynamic maritime traffic conditions
- is vital for maritime harbor supervision, safety management and collision avoidance

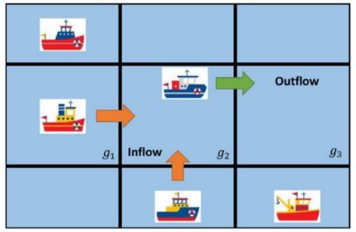
- Mandalis P., et al. (2023) **Towards a Unified Vessel Traffic Flow Forecasting Framework**. Proc. IEEE Int. Workshop BMDA.
- Mandalis P., et al. (2022) **Machine Learning Models for Vessel Traffic Flow Forecasting: An Experimental Comparison**. Proc. 3rd IEEE Int. Workshop MBDW.

VTFF and Our Contribution

- In the literature, the **most promising methods** used in predicting vessel traffic flow, mostly use **grid-based representation analysis**, which approach the VTFF problem from **two different perspectives**: a) **direct**, or b)**indirect**.

- Our work provides comparison results based on real AIS data & examines different perspectives of the VTFF problem:
  - **Direct VTFF / Sequence-based VTFF**
  - **Indirect VTFF / VRF-based VTFF**
  - **Unified Approach for VTFF (UA-VTFF)**





[2] He et al. (2017) Short-term vessel traffic flow forecasting by using an improved Kalman model. Cluster Computing.
[3] Wang et al. (2020) Use of AIS data for performance evaluation of ship traffic with speed control, Ocean Engineering.
[4] Zhou et al. (2020) Using deep learning to forecast maritime vessel flows. Sensors

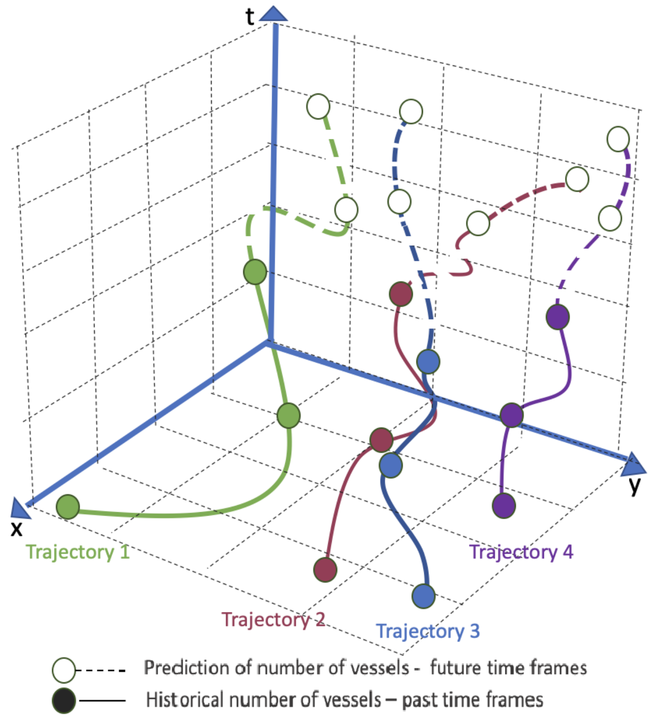
Problem Formulation

Given:

- a set of vessel trajectories $\boldsymbol{D}$ spanning in $\boldsymbol{D_s}$ (minimum bounding box of locations) space and $\boldsymbol{D_T}$ in time,
- a time duration (prediction horizon) $\boldsymbol{\Delta t}$,
- a number of temporal transitions $\boldsymbol{r}$,
- a spatiotemporal (3D) grid that partitions $\boldsymbol{D_s}$ into grid cells of resolution $\boldsymbol{G} \times \boldsymbol{G}$, and $\boldsymbol{D_T} \cup \boldsymbol{\Delta t}$ into $\boldsymbol{r}$ time frames,
- (only for UA-VTFF) a set of future vessel trajectories $\boldsymbol{D_P}$ spanning in $\boldsymbol{D_s}$ and $\boldsymbol{D_T} \cup \boldsymbol{\Delta t}$

Predict:

- The expected number of vessels (presence) in each grid cell related to $\boldsymbol{\Delta t}$.

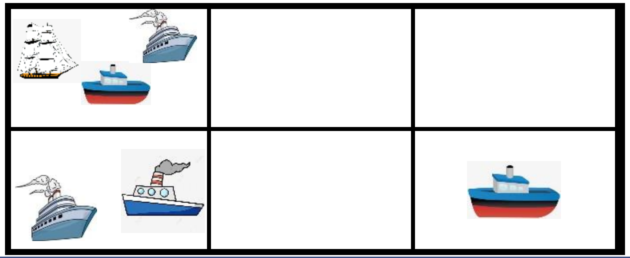


Example: Example of 4 vessel trajectories in a spatiotemporal grid of 5 time frames and 4 × 4 space resolution

Overview of our Direct & Indirect VTFF Approaches

Sequence-based VTFF

Produce a sequence of vessels' presence per spatial (2D) grid cell



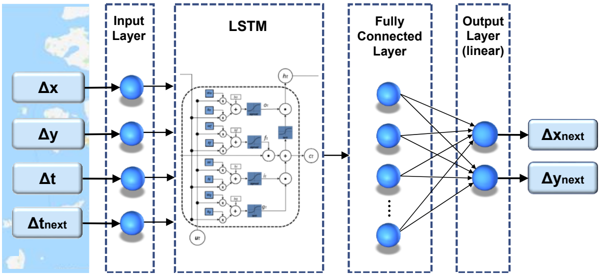
Train an ML model to predict the temporal evolution of each sequence

scikit learn

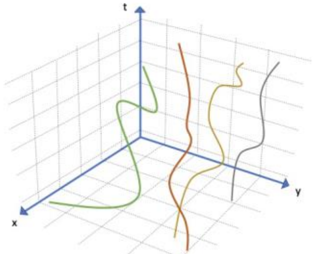
TensorFlow

VRF-based VTFF

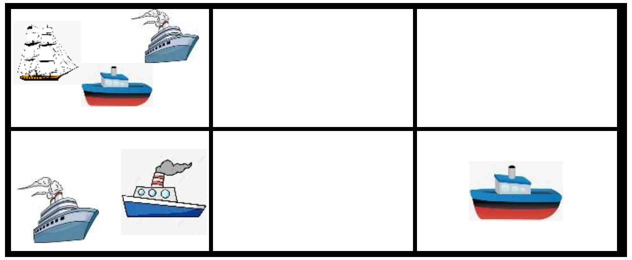
Perform VRF using a trained LSTM model



Allocate future locations into the spatio-temporal (3D) grid

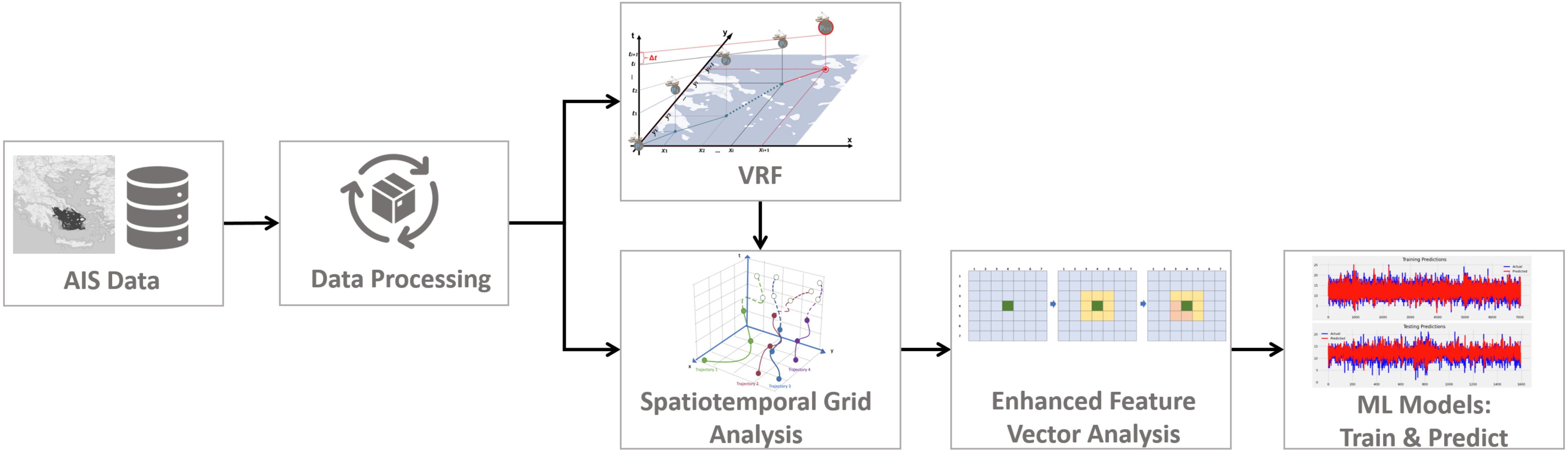


Calculate presence within each 3D grid cell



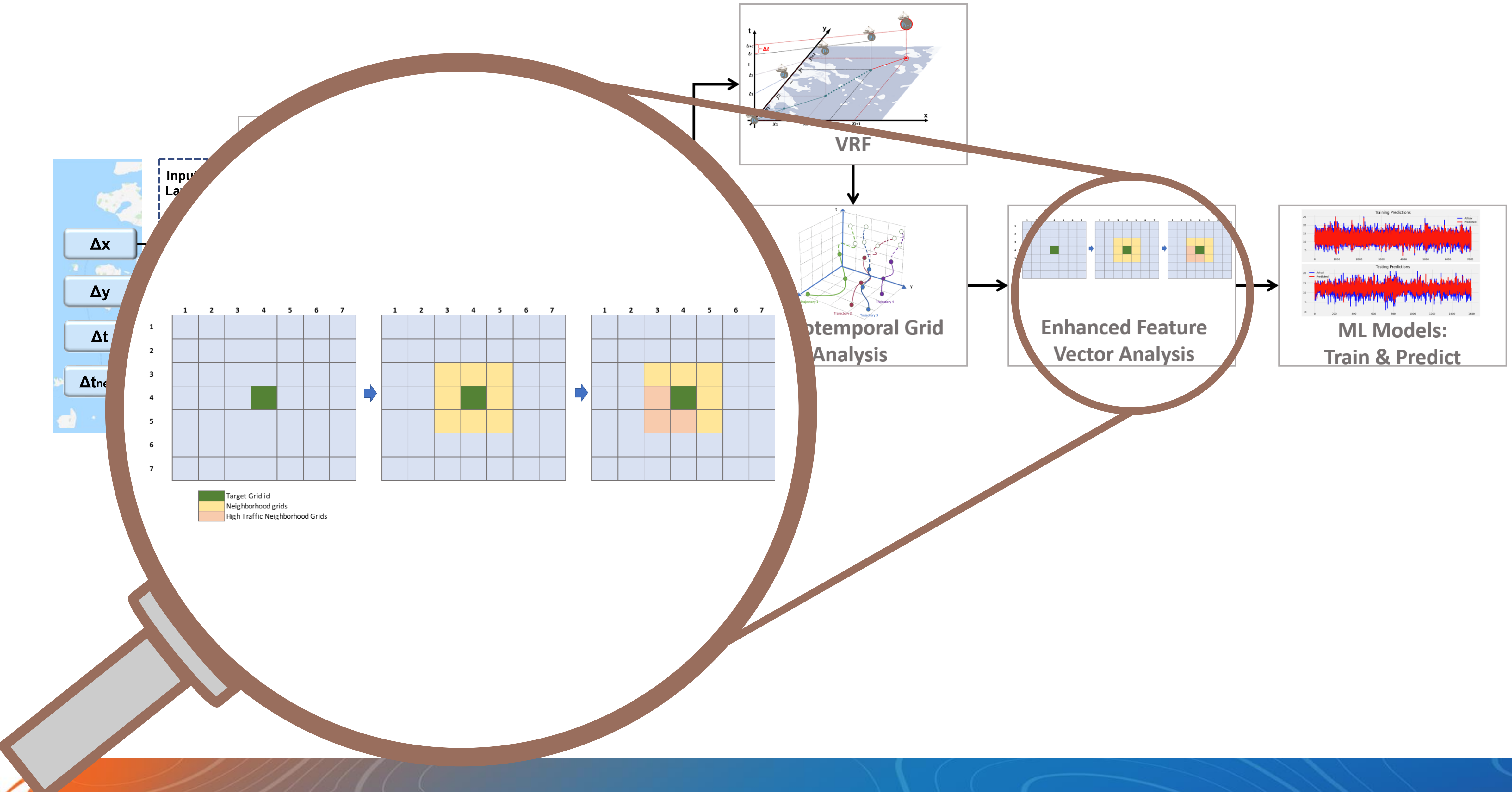
Mandalis P., et al. (2022) **Machine Learning Models for Vessel Traffic Flow Forecasting: An Experimental Comparison**. Proc. 3rd IEEE Int. Workshop MBDW.

Overview of our Unified Approach for VTFF (UA-VTFF)



Mandalis P., et al. (2023) **Towards a Unified Vessel Traffic Flow Forecasting Framework**. Proc. IEEE Int. Workshop BMDA.

Our Unified Approach for VTFF (UA-VTFF)

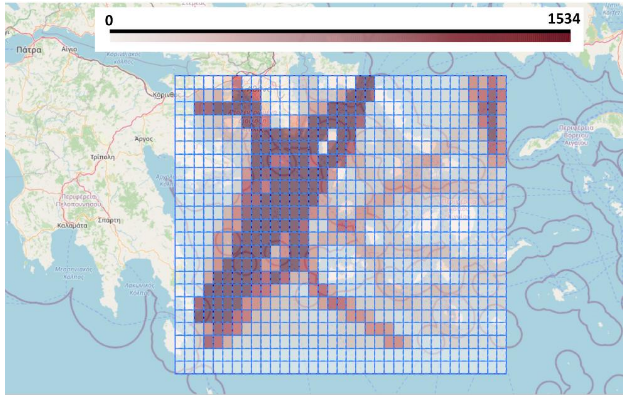


Experimental Setup

- We used a real-world dataset, called **Aegean-Cyclades**
  - 1 month (Nov. 2018) of vessel routes from/to Cyclades islands (GR)

- Experimental protocol regarding ML models:
  - Indirect VTFF / VRF-based VTFF: training (50%) | validation (25%) | testing (25%) randomly allocated
  - Direct VTFF / Sequence-based VTFF: training (initial 75%) | validation (remaining 25%) | testing (last 3 observations of the traffic flow sequence)
  - UA-VTFF: Using G=2km, Δt=30min, r=6: training (initial 75%) | validation (remaining 25%) | testing (last 6 observations of the traffic flow sequence)

- Quality measures:
  - Symmetric Mean Absolute Percentage Error (**SMAPE**)
  - **Jaccard** similarity coefficient



Overview of traffic flow (Nov. 2018), G = 10km. Darker color indicates higher traffic flow.

$$SMAPE = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{F}\sum_{t=1}^{F}2\frac{|y_{b,t}-\hat{y}_{b,t}|}{|y_{b,t}|+|\hat{y}_{b,t}|}$$

$$Jaccard = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{F}\sum_{t=1}^{F}\frac{|Y_{b,t}\cap\hat{Y}_{b,t}|}{|Y_{b,t}\cup\hat{Y}_{b,t}|}$$

Experimental Results: Direct VTFF vs Indirect VTFF

- 1st experiment: comparing the Direct & Indirect VTFF approaches (Table I)
- 2nd experiment: a closer look at the VRF-based approach (Table II)

TABLE II.
PREDICTION RESULTS (SMAPE, JACCARD) FOR THE VRF-BASED VTFF STRATEGY IN THE TESTING SET (ALL GRID CELLS) .

Grid cell (km)	Time frame (min)	SMAPE	Jaccard
5	5	9.57	0.95
	10	26.20	0.87
	15	44.00	0.78
10	5	4.97	0.97
	10	14.23	0.93
	15	24.90	0.87
15	5	3.52	0.98
	10	10.08	0.95
	15	18.04	0.91

TABLE I.
PREDICTION RESULTS (SMAPE) IN THE TESTING SET (20 BUSIEST GRID CELLS), $G = 10$KM .

VTFF strategy	Method	Time prediction horizon (min)		
		5	10	15
Flow sequence-based	XgBoost	17.72	30.41	27.43
	ARIMA	46.94	37.75	48.73
VRF-based	LSTM	6.35	16.76	28.71

Experimental Results: Unified Approach for VTFF (UA-VTFF)

Prediction results (SMAPE) for different alternatives of the UA-VTFF method in the testing set (G = 2km)

Method	Time prediction horizon (min)					
	5	10	15	20	25	30
LSTM	**12**	**16**	**27**	**26**	**27**	**26**
MLP	16	21	32	34	37	36
XgBoost	13	19	29	29	29	27
ARIMA	20	24	35	38	47	40
Prophet	21	24	33	33	38	36

Results confirm that the LSTM based UA-VTFF method can accurately capture the vessel traffic flow in short-term.

Prediction results (SMAPE) (G = 2km)

Approach	Train set	Test set
UA-VTFF (LSTM)	**18**	**21**
UA-VTFF (XgBoost)	19	22
direct VTFF [1]	23	29
indirect VTFF [1]	25	28

[1] Mandalis et.al. (2022) Machine Learning Models for Vessel Traffic Flow Forecasting: An Experimental Comparison. 23rd IEEE Int. Conf. MDM.

The proposed UA-VTFF approach (using LSTM or XgBoost) outperforms the indirect and direct VTFF strategies.

Real World Problems – Applications:
Vessel Collision Risk Assessment (VCRA)

Motivation

Vast spread of AIS-enabled maritime fleet;
Emergence of Unmanned Surface Vessels (USVs)

Motivation for several analytics (incl. **forecasting)** tasks

image source: www.ntnu.edu

Vessel Collision Risk Assessment (VCRA) is:

- critical for safety at sea
- challenging due to maritime traffic volatility
- typically addressed by calculating **Collision Risk Index** (CRI)

Tritsarolis A., et al. (2022) **Vessel Collision Risk Assessment using AIS Data: A Machine Learning Approach**. Proc. 3rd IEEE Int. Workshop MBDW.

Our Contribution vs. Related Work

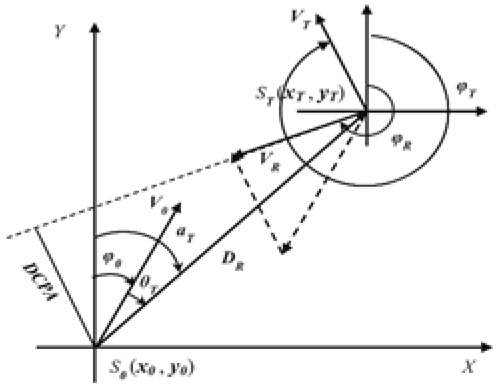
- Current state-of-the-art in VCRA → Formulaic & Deep Learning (DL) approaches
- The closest to our work combine CRI equations with…
  - Gang et al. 2016 [19]: … SVM
  - Li et al. 2018 [20]: … AFNN
  - Park et al. 2021 [21]: … RVM

- Our approach aims at decreasing processing time → investigate deeper ML architectures
  - How? by using less kinematic equations and, optionally, less features
  - Decreasing processing time → able to experiment with deeper ML architectures → yield higher accuracy & maintain the overall responsiveness of the framework.

Method	Input	Dataset(s)
SVM-VCRA [19]	$D_R, V_O, V_T, \phi_O, \phi_T, \theta_T,$	Non-public
AFNN-VCRA [20]	$D_R, V_R, \phi_R, \theta_T$	Non-public
RVM-VCRA [21]	$D_R, V_O, V_T, \phi_O, \phi_T, \theta_T, length_O, length_T$	AIS Data (Busan Port, Korea)
MLP-VCRA	$D_R,, V_O, V_T, \phi_O, \phi_T[, length_O, length_T]$	AIS Data (Saronic Gulf, Greece)

$$CRI = WU = W_{DCPA} * U_{DCPA} + W_{TCPA} * U_{TCPA} + W_D * U_D + W_B * U_B + W_K * U_K$$

$$W = [W_{DCPA}, W_{TCPA}, W_D, W_B, W_K] = [0.4457, 0.2258, 0.1408, 0.1321, 0.0556]$$

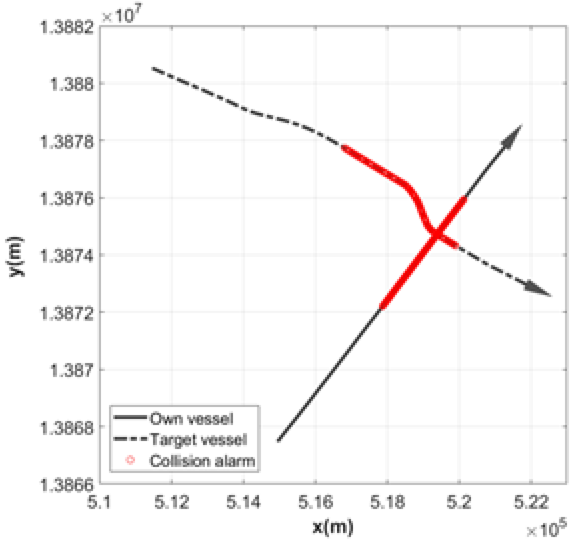


Vessel collision geometry (adapted from [19])

Problem Formulation

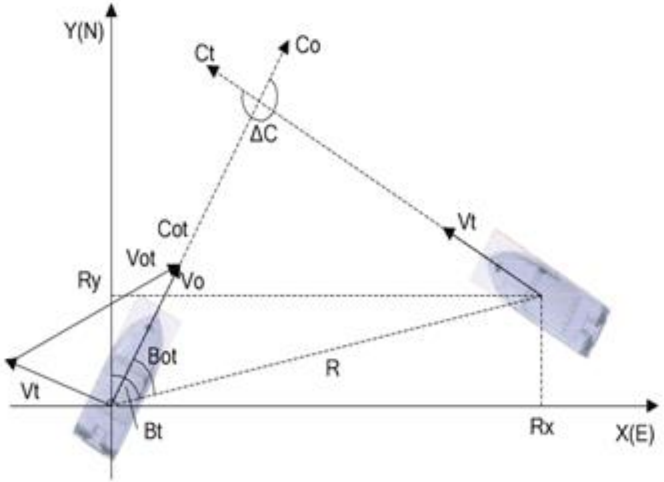
The problem: (train a ML model in order to) estimate **CRI(v_o,v_t)**, i.e., the collision risk index of an **own vessel** v_o w.r.t. a **target vessel** v_t that are in an encountering process, at **real-time**

Two vessels are in an ***encountering process*** during a time period, when their distance decreases along this time period and increases right after.



(left) Trajectories of encountering vessels in the case of crossing situation – image source: Park & Jeong 2021 [21]

(right) The moving vector diagram of encounter ships – image source: Chen et al. 2015 [7]

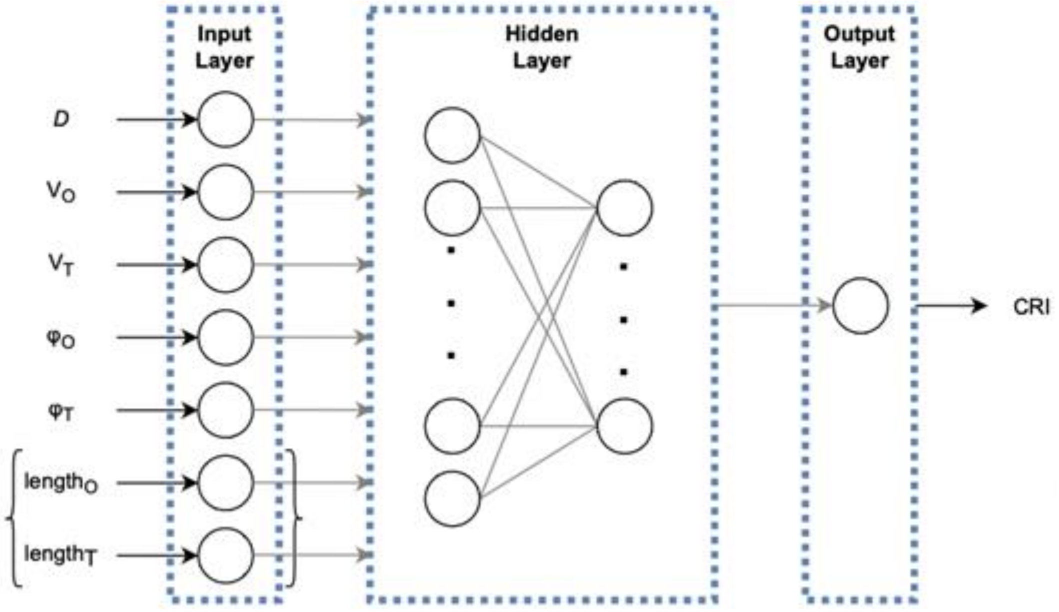


Proposed VCRA Method

Given the following features for each pair **(v_o,v_t)** of vessels in an encountering process:

- location ***(x, y)***, length, course $\boldsymbol{\varphi}$, speed ***V***

1. Create a dataset with 5+2 features:
   - distance D, speed V_O and V_T, course φ_O and φ_T
   - ❿ (optionally) $length_O$ and $length_T$
2. Train an MLP model with
   - two hidden layers (of 256 and 32 neurons, resp.)
   - one output: CRI(v_o,v_t)



The proposed MLP-VCRA architecture

Experimental Results

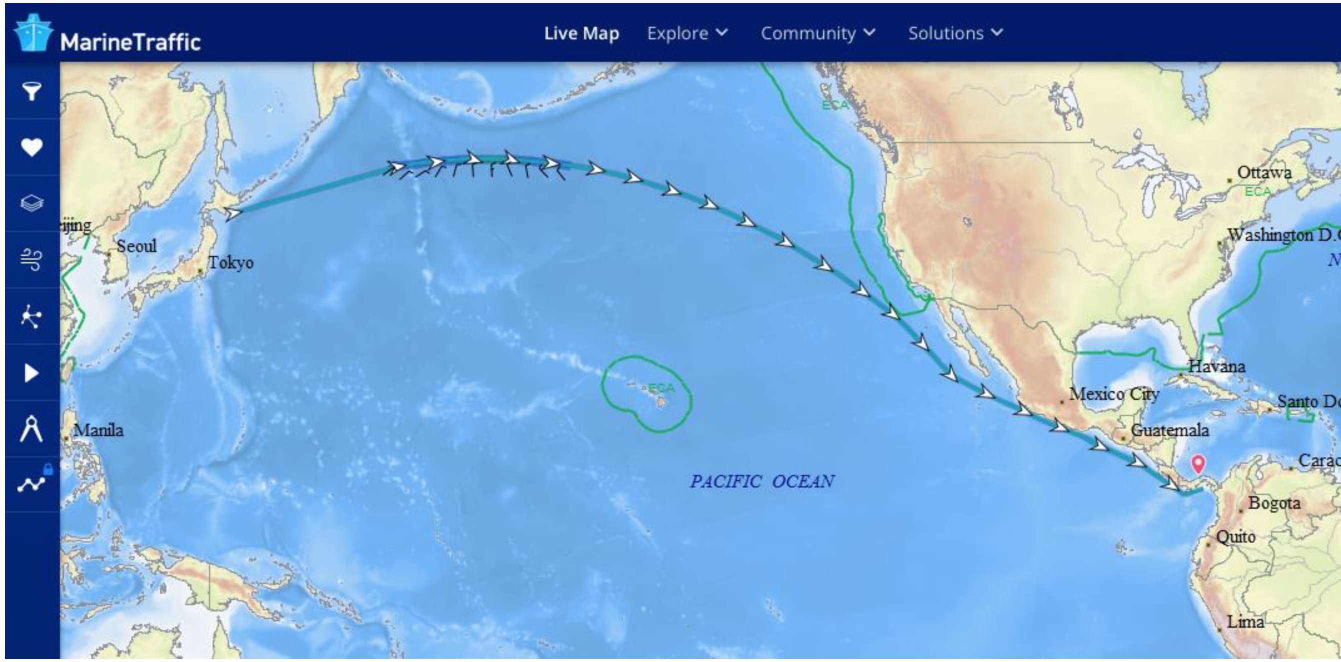
- We used a real-world dataset, called Piraeus AIS dataset
  - 1 day (July 3rd, 2018) of vessel routes in the port of Piraeus and the wider Saronic Gulf, GR
- In terms of quality, our MLP-VCRA approach
  - Reaches 87.5% accuracy after training
  - Outperforms its competitors by a large margin
- In terms of latency* (i.e., response time)
  - Outperforms competitors **and** the kinematic equations (ground truth)
- Regarding the features used
  - Vessels' length is optional & marginally improves quality and (surprisingly?) latency

Method	MAE	RMSE	Response Time (msec.)
Kinematic Eq.	-	-	329 ± 11.7
SVM-VCRA [19]	0.0572	0.0945	351 ± 1.45
AFNN-VCRA [20]	0.0476	0.0934	314 ± 2.16
RVM-VCRA [21]	0.0359	0.0802	$322 \pm .744$
MLP-VCRA	**0.0179**	**0.0485**	$\mathbf{311 \pm 1.05}$

	Accuracy (%)	MAE	RMSE	response time (msec.) (min.; med.; max.;)
MLP-VCRA ($length_O$)	86.827	0.0179	0.0485	196; 354; 680
MLP-VCRA ($length_T$)	87.134	0.0167	0.0480	201; 360; 684
MLP-VCRA ($length_{O,T}$)	**87.514**	**0.0165**	**0.0472**	**192; 332; 638**
MLP-VCRA (w/out $length_{O,T}$)	87.207	0.0189	0.0478	197; 369; 695

Summary

- **Maritime Data Analysis** & **AI** techniques can **leverage from** the "explosion" of **AIS** information in order to unlock valuable insights & pave the way for efficient Maritime Transport Systems (MTS) by tackling quite challenging tasks such as:
  - ✓ route forecasting
  - ✓ traffic forecasting
  - ✓ collision risk assessment



Acknowledgments

Research supported by EU grants:

- **MobiSpaces** – New data spaces for green mobility. 2022-25 [mobispaces.eu]
- **VesselAI** – Enabling Maritime Digitalization by Extreme-scale Analytics, AI and Digital Twins. 2021-23 [vessel-ai.eu]
- **MASTER** – Multiple Aspect Trajectory Management and Analysis, 2018-22 [master-project-h2020.eu]
- **datAcron** – Big Data Analytics for Time Critical Mobility Forecasting, 2016-18 [datacron-project.eu]
- **i4sea** – Big Data in Monitoring and Analyzing Sea Area Traffic: innovative ICT and analysis models, 2018-21 [i4sea.eu]

Thank you
for your attention!

Eva Chondrodima

evachon@unipi.gr | www.datastories.org

DATA STORIES

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

Data Science Lab.
ICT School, Univ. Piraeus

est. 2015

Check our website

www.datastories.org